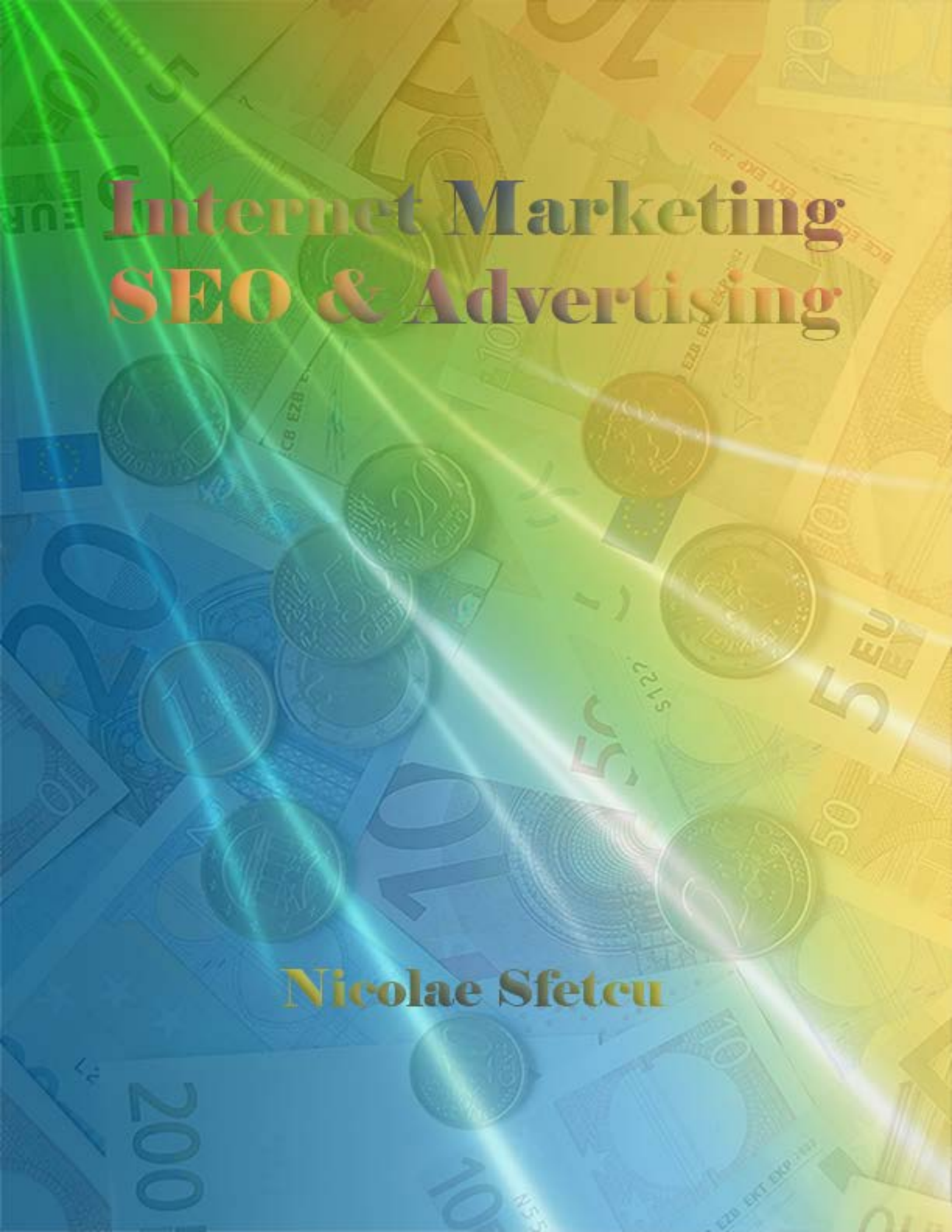

Internet Marketing SEO & Advertising

Nicolae Sfetcu

Internet Marketing, SEO & Advertising

Nicolae Sfetcu

Published by: Nicolae Sfetcu

DISCLAIMER:

The author and publisher are providing this book and its contents on an "as is" basis and make no representations or warranties of any kind with respect to this book or its contents. The author and publisher disclaim all such representations and warranties for a particular purpose. In addition, the author and publisher do not represent or warrant that the information accessible via this book is accurate, complete or current.

Except as specifically stated in this book, neither the author or publisher, nor any authors, contributors, or other representatives will be liable for damages arising out of or in connection with the use of this book. This is a comprehensive limitation of liability that applies to all damages of any kind, including (without limitation) compensatory; direct, indirect or consequential damages, including for third parties.

You understand that this book is not intended as a substitute for consultation with a licensed, educational, legal or finance professional. Before you use it in any way, you will consult a licensed professional to ensure that you are doing what's best for your situation.

This book provides content related to educational topics. As such, use of this book implies your acceptance of this disclaimer.

Contents

Advertising

Generally speaking, advertising is the promotion of goods, services, companies and ideas, usually by an identified sponsor. Marketers see advertising as part of an overall promotional strategy. Other components of the promotional mix include publicity, public relations, personal selling and sales promotion.

History

In ancient times the most common form of advertising was "word of mouth". However, commercial messages and election campaign displays were found in the ruins of Pompeii. Egyptians used papyrus to create sales messages and wall posters. Lost-and-found advertising on papyrus was common in Greece and Rome. Wall or rock painting for commercial advertising is another manifestation of an ancient media advertising form which is present to this day in many parts of Asia, Africa, and South America. For instance, tradition of wall paintings may be traced back to India rock-art paintings that goes back to 4000 BC, see Bhatia 2000: 62-68 on the evolution of wall advertising. As printing developed in the 15th and 16th century, advertising expanded to include handbills. In the 17th century advertisements started to appear in weekly newspapers in England.

These early print ads were used mainly to promote books (which were increasingly affordable) and medicines (which were increasingly sought after as disease ravaged Europe). Quack ads became a problem, which ushered in regulation of advertising content.

As the economy was expanding during the 19th century, the need for advertising grew at the same pace. In America, the classified ads became popular, filling pages of newspapers with small print messages promoting all kinds of goods. The success of this advertising format led to the growth of mail-order advertising. In 1843 the first advertising agency was established by Volney Palmer in Philadelphia. At first the agencies were just brokers for ad space in newspapers, but by the 20th century, advertising agencies started to take over responsibility for the content as well.

The 1960s saw advertising transform into a modern, more scientific approach in which creativity was allowed to shine, producing unexpected messages that made advertisements interesting to read. The Volkswagen ad campaign featuring such headlines as "Think Small" and "Lemon" ushered in the era of modern advertising by promoting a "position" or "unique selling proposition" designed to associate each brand with a specific idea in the reader or viewer's mind.

Today, advertising is evolving even further, with "guerrilla" promotions that involve unusual approaches such as staged encounters in public places, giveaways of products such as cars that are covered with brand messages, and interactive advertising where the viewer can respond to become part of the advertising message.

Media

Commercial advertising media can include wall paintings, billboards (outdoor advertising), street furniture components, printed flyers, radio, cinema and television ads, web banners, web popups, skywriting, bus stop benches, magazines, newspapers, town criers, sides of buses, taxicab doors and roof mounts, musical stage shows, subway platforms and trains, elastic bands on disposable diapers, stickers on apples in supermarkets, the opening section of streaming audio and video, and the backs of event tickets and supermarket receipts. Any place an "identified" sponsor pays to deliver their message through a medium is advertising.

Covert advertising embedded in other entertainment media is known as product placement. A more recent version of this is advertising in film, by having a main character use an item or other of a definite brand - an example is in the movie *Minority Report*, where Tom Cruise's character Tom Anderton owns a computer with the *Nokia* logo clearly written in the top corner, or his watch engraved with the *Bulgari* logo. Another example of advertising in film is in *I, Robot*, where main character played by Will Smith mentions his *Converse* shoes several times, calling them "classics," because the film is set far in the future.

The TV commercial is generally considered the most effective mass-market advertising format and this is reflected by the high prices TV networks charge for commercial airtime during popular TV events. The annual Super Bowl football game in the United States is known as much for its commercial advertisements as for the game itself, and the average cost of a single thirty-second TV spot during this game has reached $2.5 million (as of 2006).

Virtual advertisements may be inserted into regular television programming through computer graphics. It is typically inserted into otherwise blank backdrops or used to replace local billboards that are not relevant to the remote broadcast audience. More controversially, virtual billboards may be inserted into the background where none existing in real-life. Virtual product placement is also possible.

Increasingly, other mediums such as those discussed below are overtaking television due to a shift towards consumer's usage of the Internet as well as devices such as TiVo.

Advertising on the World Wide Web is a recent phenomenon. Prices of Web-based advertising space are dependent on the "relevance" of the surrounding web content and the traffic that the website receives.

E-mail advertising is another recent phenomenon. Unsolicited bulk E-mail advertising is known as "spam".

Some companies have proposed to place messages or corporate logos on the side of booster rockets and the International Space Station. Controversy exists on the effectiveness of subliminal advertising, and the pervasiveness of mass messages.

Unpaid advertising (also called word of mouth advertising), can provide good exposure at minimal cost. Personal recommendations ("bring a friend", "sell it by zealot"), spreading buzz, or achieving the feat of equating a brand with a common noun ("Xerox" = "photocopier", "Kleenex" = tissue, "Vaseline" = petroleum jelly, "Kotex" = tampons, "Maxi

pads" = sanitary napkins, "Scotch Tape" = Clear Tape, "Band-aid" = bandage, "Visine" = eye drops, "Q-tips" = cotton swabs, "Rollerblades" = inline skates) -- these must provide the stuff of fantasy to the holder of an advertising budget.

The most common method for measuring the impact of mass media advertising is the use of the rating point (rp) or the more accurate target rating point (trp). These two measures refer to the percentage of the universe of the existing base of audience members that can be reached by the use of each media outlet in a particular moment in time. The difference between the two is that the rating point refers to the percentage to the entire universe while the target rating point refers to the percentage to a particular segment or target. This becomes very useful when focusing advertising efforts on a particular group of people. For example, think of an advertising campaign targeting a female audience aged 25 to 45. While the overall rating of a TV show might be well over 10 rating points it might very well happen that the same show in the same moment of time is generating only 2.5 trps (being the target: women 25-45). This would mean that while the show has a large universe of viewers it is not necessarily reaching a large universe of women in the ages of 25 to 45 making it a less desirable location to place an ad for an advertiser looking for this particular demographic.

Impact

"Half the money I spend on advertising is wasted; the trouble is, I don't know which half." - John Wanamaker, father of modern advertising.

The impact of advertising has been a matter of considerable debate and many different claims have been made in different contexts. During debates about the banning of cigarette adervertising, a common claim from cigarette manufacturers was that cigarette advertising does not encourage people to smoke who would not otherwise. The (eventually successful) opponents of advertising, on the other hand, claim that advertising does in fact increase consumption

According to many media sources, the past experience and state of mind of the person subjected to advertising may determine the impact that advertising has. Children under the age of four may be unable to distinguish advertising from other television programs, whilst the ability to determine the truthfullness of the message may not be developed until the age of eight.

Public service advertising

The same advertising techniques used to promote commercial goods and services can be used to inform, educate and motivate the public about non-commercial issues, such as AIDS, political ideology, energy conservation, religious recruitment, and deforestation.

Advertising, in its non-commercial guise, is a powerful educational tool capable of reaching and motivating large audiences. "Advertising justifies its existence when used in the public interest - it is much too powerful a tool to use solely for commercial purposes." - Attributed to Howard Gossage by David Ogilvy

Public service advertising, non-commercial advertising, public interest advertising, cause marketing, and social marketing are different terms for (or aspects of) the use of

sophisticated advertising and marketing communications techniques (generally associated with commercial enterprise) on behalf of non-commercial, public interest issues and initiatives.

In the United States, the granting of television and radio licenses by the FCC is contingent upon the station broadcasting a certain amount of public service advertising. To meet these requirements, many broadcast stations in America air the bulk of their required Public Service Announcements during the late night or early morning when the smallest percentage of viewers are watching, leaving more day and prime time commercial slots available for high-paying advertisers.

Public service advertising reached its height during World Wars I and II under the direction of several U.S. government agencies.

Flyposting

Flyposting is the act of placing advertising posters or flyers in illegal places. In the US, these posters are known as **bandit signs** or **street spam**.

In most areas, it is illegal to place such posters on private property without the consent of the property owner or on public property without a sign permit from the local government.

It is an advertising tactic mostly used by small businesses promoting concerts and political activist groups, but there have been occasions where international companies subcontracted local advertising agencies for flyposting jobs in order to not get caught in illegal behavior.

Flyposting is commonly seen as a nuisance due to issues with property rights, visual appearance and littering and is a misdemeanor in many countries.

Social impact

Regulation

There have been increasing efforts to protect the public interest by regulating the content and the reach of advertising. Some examples are the ban on television tobacco advertising imposed in many countries, and the total ban on advertising to children under twelve imposed by the Swedish government in 1991. Though that regulation continues in effect for broadcasts originating within the country, it has been weakened by the European Court of Justice, which has found that Sweden was obliged to accept whatever programming was targeted at it from neighbouring countries or via satellite.

In Europe and elsewhere there is a vigorous debate on whether and how much advertising to children should be regulated. This debate was exacerbated by a report released by the Henry J. Kaiser Family Foundation in February 2004 which suggested that food advertising targeting children was an important factor in the epidemic of childhood obesity raging across the United States.

In many countries - namely New Zealand, South Africa, Canada, and many European contries- the advertising industry operates a system of self-regulation. Advertisers, advertising agencies and the media agree on a code of advertising standards that they

attempt to uphold. The general aim of such codes is to ensure that any advertising is 'legal, decent, honest and truthful'. Some self-regulatory organisations are funded by the industry, but remain independent, with the intent of upholding the standards or codes (like the ASA in the UK).

Critiques of the medium

As advertising and marketing efforts become increasingly ubiquitous in modern Western societies, the industry has come under criticism of groups such as AdBusters via culture jamming which criticizes the media and consumerism using advertising's own techniques. The industry is accused of being one of the engines powering a convoluted economic mass production system which promotes consumption. Some advertising campaigns have also been criticized as inadvertently or even intentionally promoting sexism, racism, and ageism. Such criticisms have raised questions about whether this medium is creating or reflecting cultural trends. At very least, advertising often reinforces stereotypes by drawing on recognizable "types" in order to tell stories in a single image or 30 second time frame. Recognizing the social impact of advertising, MediaWatch, a non-profit women's organization, works to educate consumers about how they can register their concerns with advertisers and regulators. It has developed educational materials for use in schools. The award-winning book, Made You Look - How Advertising Works and Why You Should Know , by former MediaWatch president Shari Graydon, provides context for these issues for young readers.

Public interest groups and free thinkers are increasingly suggesting that access to the mental space targeted by advertisers should be taxed, in that at the present moment that space is being freely taken advantage of by advertisers with no compensation paid to the members of the public who are thus being intruded upon. This kind of tax would be a Pigovian tax in that it would act to reduce what is now increasingly seen as a public nuisance. Efforts to that end are gathering momentum, with Arkansas and Maine considering bills to implement such taxation. Florida enacted such a tax in 1987 but was forced to repeal it after six months, as a result of a concerted effort by national commercial interests, which withdrew planned conventions, causing major losses to the tourism industry, and cancelled advertising, causing a loss of 12 million dollars to the broadcast industry alone.

Public perception of the medium

Over the years, the public perception of advertising has become very negative. It is seen as a medium that inherently promotes a lie, based on the purpose of the advertisement - to encourage the target audience to submit to a cause or a belief, and act on it to the advertising party's benefit and consequently the target's disadvantage. They are either perceived as directly lying (stating opinions or untruths directly as facts), lying by omission (usually of terms unfavorable to the customer), portraying a product or service in a light that does not reflect reality or even making up realities where their product has a new role. Yet as with many other things in life, the vast majority of the public do not care enough to act. One can either choose to listen to the many campaigns or to ignore them.

Effects on communication media

Another effect of advertising is to modify the nature of the communication media where it is shown. The clearest example is television. Channels that get most of their revenues from publicity try to make their medium a good place for communicating ads. That means trying to make the public stay for long times and in a mental state that will make spectators not to switch the channel through the ads. Programs that are low in mental stimulus and require light concentration and are varied are best for long sitting times and make for much easier emotional jumps to ads, that can become more entertaining than regular shows. A simple way to understand the objectives in television programming is to compare contents from channels paid and chosen by the viewer with channels that get their income mainly from advertisements.

Future

With the dawn of the Internet have come many new advertising opportunities. Popup, Flash, banner, and email advertisements (the last often being a form of spam) abound. Recently, the advertising community has attempted to make the adverts themselves desirable to the public. In one example, Cadillac chose to advertise in the movie 'The Matrix Reloaded', which as a result contained many scenes in which Cadillac cars were used. Similarly, product placement for Rolex watches and BMW cars featured in recent James Bond films.

Each year, greater sums are paid to obtain a commercial spot during the Super Bowl. Companies attempt to make these commercials sufficiently entertaining that members of the public will actually want to watch them.

Particularly since the rise of "entertaining" advertising, some people may like an advert enough that they wish to watch it later or show a friend. In general, the advertising community has not yet made this easy, although some have used the Internet to widely distribute their adverts to anyone wishing to see or hear them.

Advertising agencies

An **advertising agency** or **ad agency** is a service business dedicated to creating, planning and handling advertising (and sometimes other forms of promotion) for their clients. An ad agency is independent from the client and provides an outside point of view to the effort of selling the client's products or services. An agency can also handle overall marketing and branding strategies and sales promotions for its clients.

Typical ad agency clients include businesses and corporations, non-profit organizations and government agencies. Agencies may be hired to produce single ads or, more commonly, ongoing series of related ads, called an advertising campaign.

Ad agencies come in all sizes, from small one- or two-person shops to large multi-national, multi-agency conglomerates such as Omnicom Group or WPP Group.

Some agencies specialize in particular types of advertising, such as print ads or television commercials. Other agencies, especially larger ones, produce work for many types of media

(creating integrated marketing communications, or through-the-line (TTL) advertising). The "line", in this case, is the traditional marker between media that pay a (traditionally 15%) commission to the agency (mainly broadcast media) and the media that do not.

Lately, Search Engine Marketing (SEM) and Search Engine Optimization (SEO) firms have been classified by some as 'agencies' due to the fact that they are creating media and implementing media purchases of text based (or image based in some instances of search marketing) ads. This relatively young industry has been slow to adopt the term 'agency' however with the creation of ads (either text or image) and media purchases they do qualify technically as an 'advertising agency' as well as recent studies suggest that both SEO and SEM are set to outpace magazine spending in the next 3-5 years.

Not all advertising is created by agencies. Companies that create and plan their own advertising are said to do their work *in house*.

Agency personnel

The *creative department* -- the people who create the actual ads -- form the core of an advertising agency. Modern advertising agencies usually form their copywriters and art directors into creative teams. Creative teams may be permanent partnerships or formed on a project-by-project basis. The art director and copywriter report to a creative director, usually a creative employee with several years of experience. Although copywriters have the word "write" in their job title, and art directors have the word "art", one does not necessarily write the words and the other draw the pictures; they both generate creative ideas to represent the proposition (the advertisement or campaign's key message).

The other major department in ad agencies is *account services* or *account management*. Account service employees work directly with clients and potential clients, soliciting business for the ad agency and determining what clients need and want the agency to do for them. They are also charged with understanding the clients business situation and representing those needs within the agency, so that ads can be brought to bear on the correct problem.

Previously, client services employees wrote the advertising strategy that the creative director (and teams) would use to create the advertising. However, since the late 1960's in the UK, and the mid-1980's in the US, specialist account planners have been tasked with doing this. The account planner was originally employed to "represent the consumer" in the advertising i.e. find the best way to pitch the clients products to people but better understanding them, what they want and how to talk to them. Planning's role has expanded considerably since it was originally introduced. Pleanners now brand strategists and, to a certain extent, media strategists - using consumer insights to understand where and how people are most receptive to certain messages.

The *creative services* department may not be so well known, but its employees are the people who have contacts with the suppliers of various creative media. For example, they will be able to advise upon and negotiate with printers if an agency is producing flyers for a client. However, when dealing with the major media (broadcast media, outdoor, and the press), this work is usually outsourced to a media agency which can advise on *media*

planning and is normally large enough to negotiate prices down further than a single agency or client can.

In small agencies, employees may do both creative and account service work. Larger agencies attract people who specialize in one or the other, and indeed include a number of people in specialized positions: production work, [Internet] advertising, or research, for example.

An often forgotten, but extremely important, department within an advertising agency is traffic. Typically headed by a traffic manager (or system administrator), this department is responsible for a number of things. First and foremost is increasing agency efficiency and profitability through the reduction of false job starts, inappropriate job initiation, incomplete information sharing, over- and under-cost estimation, and the need for media extensions. In small agencies without a dedicated traffic manager, one employee may be responsible for managing workflow, gathering cost estimates and answering the phone, for example. Large agencies may have a traffic department of ten or more employees. Department size varies, but its importance remains the same.

Online advertising

Online advertising is advertising on the Internet. This particular form of advertising is a source of revenue for an increasing number of websites and companies.

There are two sides to online advertising, a legitimate one and an illegitimate one. The legitimate side of online advertising includes search engine advertising, advertising networks and opt-in e-mail advertising. The illegitimate side is dominated by spamming.

Though the range of advertising options has expanded since in the commercialization of the Internet, the use of rich media and static images is extremely popular. The ever-increasing audience of online users will likely continue to be a major advertising market.

Overview of the market

A significant number of firms, from small businesses to multinational corporations, incorporate online advertising into their marketing strategy. This is even true of firms which conduct their business through more traditional brick and mortar channels. In response to this demand, a number of firms specialize in facilitating online marketing. Therefore, online advertisements typically involve at least two separate firms: the advertiser or agency which purchases or sponsors the advertisement and the publisher or network which distributes the ad for display. Additional parties may also be included, such as an ad serving technology provider, a third party sales network, or other combinations.

In capitalizing on the increasing importance of the Internet as a marketing medium, the online advertising industry has developed specialized technical systems to manage the ways ads are distributed and viewership totaled. The Internet Advertising Bureau (IAB) has established guidelines for the counting methodology, size requirements, and other aspects of the business.

Because of the close relation between technical innovation and online advertising, many firms specialize in both. For example, most search engines couple their search service with an advertising program, exploiting the benefits of keyword-based search technology by including ads in search results. Many technology firms specialize in ad serving, the systems used to select the ads to show, optimize results, and generate reports.

Payment conventions

Because of the ability to track results of online advertising at a more granular level than what is available through traditional advertising, varying ways have developed for the advertisers and publishers to do business. The three most common ways in which online advertising is purchased are CPA, CPC, and CPM.

CPA (Cost Per Action) advertising is performance based and is common in the affiliate marketing sector of the business. In this payment scheme, the publisher takes all the risk of running the ad, and the advertiser only pays for the media on the basis of the number of users who complete a transaction, such as a purchase or sign-up.

CPC (Cost Per Click) advertising is also performance based and is common in search marketing, where it is often known as Pay per click (PPC). In this scheme, an advertisement may be displayed (and assumedly viewed) many times, but the advertiser only pays based on the number of user clicks. This system provides an incentive for publishers to target ads correctly (often by keyword), as the payment depends upon the ad not only being seen, but the viewer responding and following the hyperlink.

CPM (Cost per Thousand) advertising is the most common basis in the business and is used for most display advertising and rich media. This scheme most closely resembles offline advertising, wherein the advertiser is paying for exposure of their message to a specific audience. CPM costs are priced per thousand, so that a $1 CPM, means that the advertiser pays $1 for every thousand impressions.

Rich Media advertising

The display advertising portion of online advertising is increasingly dominated by rich media, generally using Macromedia Flash. Rich media advertising techniques make overt use of color, imagery, page layout, and other elements in order to attract the reader's attention. Some users might consider these ads as intrusive or obnoxious, because they can distract from the desired content of a webpage. Some examples of common rich media formats and the terms of art used within the industry to describe them:

- Interstitial or Expanding ad: The display of a page of ads before the requested content.
- Floating ad: An ad which moves across the user's screen or floats above the content.
- Expanding ad: An ad which changes size and which may alter the contents of the webpage.
- Polite ad or Polite download: A method by which a large ad will be downloaded in smaller pieces to minimize the disruption of the content being viewed
- Wallpaper ad: An ad which changes the background of the page being viewed.

In addition, ads containing streaming video or streaming audio are becoming very popular with advertisers.

Email advertising

Legitimate Email advertising is often known as opt-in email to distinguish it from spam.

Affiliate marketing

Affiliate marketing is a form of advertising where the advertiser allows a potentially large number of small publishers to pick specific creative elements or offers to market in exchange for payment should such marketing create sales or other revenue. This is usually accomplished though a self-service online system, such as those offered by third parties Performics, BeFree, CommissionJunction, or Linkshare. Affiliate marketing was an early innovation of online retailer Amazon, which has used its program to generate enormous volumes of low cost brand exposure.

Contextual advertising

Contextual advertising is the term applied to advertisements appearing on websites which ads are served by automated systems based on the content of the page.

Google AdSense was the first major contextual advertising program. It worked by providing webmasters with JavaScript code that, when inserted into web pages, called up relevant advertisements from the Google inventory of advertisers. The relevance was calculated by a separate Google bot that indexed the content of the page.

Since the advent of AdSense, the Yahoo Publisher Network, MSN adCenter and others have been gearing up to make similar offering.

Contextual advertising has made a major impact on earnings of many websites. As the ads are more targeted they are more likely to get clicked, thus generating revenue for the owner of the website (and the server of the advertisement). A large part of Google's earnings are from their share of the contextual ads served on the millions of webpages running the Adsense program.

Many advertising networks display text-only ads that correspond to the keywords of an Internet search or to the content of the page on which the ad is shown. These ads are believed to have a greater chance of attracting a user, because they tend to share a similar context as the user's search query. For example, a search query for "flowers" might return an advertisement for a florist's website.

Another newer technique is embedding keyword hyperlinks in a webpage which are sponsored by an advertiser. When a user follows the link, they are sent to a sponsor's website.

Domain parking

Domain parking is an advertising practice used primarily by domain name registrars and internet advertising publishers to monetize type-in traffic visiting an inactive domain name. The undeveloped domain name will usually point (redirect) to a page full of advertising links. These links will be targeted to the predicted interests of the visitor. Usually the domain owner is paid based on how many links have been visited (e.g. pay per click) or how beneficial they have been. The keywords of the domain name provide clues as to what the visitor sought before arriving. For example, the domain fast-weight-loss.com might point to a page of links to fad diets and drugs.

Another use of domain parking is to be a placeholder of an existing web site. A company might choose to use this method to redirect its web site traffic to another web site it owns.

Type-in traffic

Type-in traffic is a term describing visitors landing at a web site by entering a word or phrase in the web browser's address bar rather than following a hyperlink from another web page, using a browser bookmark, or a search-box search. For example, if you are interested in widgets, then instead of doing a search for 'widgets' you might type 'widgets.com' in your address bar to see if such a web site exists, and, if so, what content is there. From another perspective, if you are in the business of selling widgets, then owning

the domain name 'widgets.com' and having an active website at that address would be a desirable thing, as you could take advantage of the targeted type-in traffic that this name receives. That simple example holds true for virtually all products and services.

Most web browsers formerly defaulted the top-level domain to *com*, thus entering 'mysearchterm' in the web browser's address bar usually would lead the user to *http://mysearchterm.com/*. This behavior changed depending on the 'default search engine' setting in the web browser's properties, so entering 'mysearchterm' in the address bar could also lead to an error page or to results from a search engine. Today most error page traffic has been taken over by browser manufacturers such as Microsoft and Netscape for the purpose of displaying paid search advertising. Much of MSN's high ranking as a portal results from the error page traffic delivered from their dominant Internet Explorer browsers.

In the last few years advertisers, publishers and ad networks such as MSN, AOL, Google and Yahoo have awoken to the power of displaying relevant advertising to highly targeted type-in traffic from domain names, browser address bar searches and error traffic.

In November 2004 Marchex acquire d the generic domain name portfolio of Name Development Ltd., a little known British Virgin Islands company, for 164 million dollars, predominantly for its 100,000+ domain name portfolio. This portfolio generated 17 million type-in traffic visitors each month.

In 2005, Highland Capital, a venture capital firm, acquired a controlling interest in BuyDomains, paying an undisclosed sum for its domain name portfolio.

In August 2005, industry trade journals such as dnjournal, dnforum and domainstate reported that sale volumes and prices of existing generic domain-names were rising rapidly as a result of type-in traffic monetization opportunities. Small webmasters can buy a domain name with type-in traffic and utilize Google's Adsense product, or any of several traffic aggregators such as Namedrive, Fabulous, DomainSponsor, or Skenzo to display relevant advertising to the trickle of visitors coming to their domain names. Many small publishers are generating thousands of dollars each month in revenue with very little effort by building websites that serve relevant advertising to their type-in traffic visitors.

Google's entry into the small publisher monetization space came as a result of their purchase of Applied Semantics in 2003. The drop registrar phenomenon is directly related to the value and desireability of type-in traffic domain names.

Type-in traffic does not differentiate between trademark traffic and generic traffic as it relates to domain names. For example, the act of registering coca-cola.com for one's own commercial gain would be considered cybersquatting. However, the act of registering softdrinks.com or cola.com would likely be a defensible acquisition of a generic domain-name for type-in traffic generation or resale business opportunities.

Advertising networks

An **advertising network** (also called an **online advertising network** or **ad network**) is a collection of (often unrelated) online advertising inventory.

Online advertising inventory comes in many different forms. This inventory can be found on websites, in instant messaging applications, in adware, in e-mails, and on other sources. Some examples of advertising inventory include: banner ads, rich media, text links, and e-mails. (This is not an exhaustive list.)

Large publishers often sell only their remnant inventory through ad networks. While not commonly known, even among many large publishers remnant inventory can exceed 50% of total inventory, although this is not always the case. Typical numbers range from 10% to 60% of total inventory being remnant and sold through advertising networks.

Smaller publishers often sell all of their inventory through ad networks. One type of *ad network*, know as the *blind network*, is such that advertisers place ads, but do not know the exact places where their ads are being placed.

In most cases, ad networks deliver their content through the use of a central ad server.

Classified ads

Classified advertising is a form of advertising which is particularly common in newspapers and other periodicals. Classified advertising is usually textually based and can consist of as little as the type of item being sold, (i.e., "Clothing") and a telephone number to call for more information ("call 555-7777"). It can also have much more detail, such as name to contact, address to contact or visit, a detailed description of the product or products ("pants and sweaters, size 10" as opposed to "clothing", "red 1996 Pontiac Grand Prix" as opposed to "automobile"). There are generally no pictures or other graphics within the advertisement, although sometimes a logo may be used. Classified advertising is called such because it is generally grouped within the publication under headings classifying the product or service being offered (headings such as Accounting, Automobiles, Clothing, Farm Produce, For Sale, For Rent, etc.) and is grouped entirely in a distinct section of the periodical, which makes it distinct from display advertising, which often contains graphics or other art work and which is more typically distributed throughout a publication adjacent to editorial content. A hybrid of the two forms — **classified display advertising** — may often be found, in which categorized advertisements with larger amounts of graphical detail can be found among the text listings of a classified advertising section in a publication. Business opportunities often use classifieds to sell their services, usually employing 1-800 numbers. Classified ads are also among the tools used by many companies in recruitment for available job opportunities.

In recent years the term "classified advertising" or "classified ads" has expanded from merely the sense of print advertisements in periodicals to include similar types of advertising on computer services, radio, and even television, particularly cable television but occasionally broadcast television as well, typically very early in the morning hours.

Like most forms of printed media, the classified ad has found its way to the Internet. Printed classified ads are typically just a few column lines in length, and they often filled with abbreviations to save space and money. Internet classified ads do not typically use per-line pricing models, so they tend to be longer. They are also more readily searchable unlike their offline brethren, and tend to be local classifieds with a great sense of urgency because of their daily structure. Because of their self-policing nature and low cost

structures, some companies offer free online classified ads such as Craigslist, Lazycity, Classified.Ad and AdPost. Craigslist was one of the first online classified sites, and is currently one of the largest. There are also country-specific classified sites like Bechna.com in India or Gumtree from the UK. There are a number of agencies throughout the world that have made a business out of the classified advertising industry. For example Wide Area Classifieds has created a classified network where people can place ads in papers across the US.

In 2003, the market for classified ads in the United States was $15.9 billion (newspapers), $14.1 billion (online) according to market researcher Classified Intelligence. The worldwide market for classified ads in 2003 was estimated at over $100 billion.

As the online classified advertising sector develops, there is an increasing emphasis toward specialisation. Like search engines, classified websites are often vertical in nature with sites providing advertising platforms for niche markets of buyers of sellers.

Ad serving

Ad serving describes the technology and service that places advertisements on web sites. Ad serving technology companies provide software to web sites and advertisers to serve ads, count them, choose the ads that will make the web site or advertiser most money, and monitor progress of different advertising campaigns.

Two types of internet companies use ad serving: web sites and advertisers. The main purpose of using an ad server is different for both of them:

For a **web site**, the ad server needs to look through all the ads available to serve to a user who is on a page, and choose the one that will make the web site the most money, but still conform to the rules that the advertiser and web site have agreed. For example if a web site has 10 different advertisers that have paid for a big square ad, the ad server must decide which one to serve (or display). One advertiser may have only agreed to pay for ads from 9am - 5pm. If it is after 5pm, then the Ad Server must not serve that one. Another advertiser may only have paid to show one ad to each user per day. The ad server must therefore see if a user has seen that ad before, on that day and not serve it again if the user has seen it. Another advertiser may have agreed a high price, but only if the person watching the page is in the United States. In that case, the Ad Server needs to check the IP address to determine if the user is in the US and then decide which is the highest paying ad for that user, in the US, at that time, given what that user has seen in the past.

For an **advertiser** the ad server needs to try to serve the ad that is most likely to result in a sale of the product advertised. For example if a user is viewing a page, the advertiser's ad server needs to decide from previous history, what ad that user is most likely to click on and then buy the product advertised. If the user is on a technology page, then the ad server may know that on technology types of pages, the ad that works best is a blue one with mostly text and pricing and numbers, not the green ad with a picture of a model and little text. The ad server will therefore serve this ad, to try and get the highest probability of a sale from the ad.

Ad Serving is most complex when it is used by an Advertising Network. An advertising network buys ads from many web sites and therefore acts like an advertiser user of Ad Serving. When the network buys ads, it tries to place ads on sites where they work best. However an ad network then sells its aggregated ad inventory to advertisers. When doing this, it uses its Ad Serving software as a web site does. In this case it tries to make the most money by only running the ads from advertisers that pay most.

Central ad server

A **central ad server** is a computer server that stores advertisements and delivers them to web site visitors. These servers centrally store the ads so that advertisers and publishers can track from one source the distribution of their online advertisements, and have one location for controlling the rotation and distribution of their advertisements across the web.

The central ad server was first developed and introduced by FocaLink Media Services in 1995 for controlling online advertising or banner ads. The company was founded by Dave Zinman and Jason Strober, and based in Palo Alto, CA. In 1998, the company changed its name to AdKnowledge, and was eventually purchased by CMGI in 1999.

Ad Server Functionality

The typical common functionality of ad servers includes:

1. Uploading creative, including display advertisements and rich media
2. Trafficking ads according to differing business rules
3. Targeting ads to different users, or content
4. Optimizing creative based on results
5. Reporting impressions, clicks, post-click activities, and interaction metrics

Advanced functionality may include:

- Frequency capping creative so users only see messages a limited amount of time
- Sequencing creative so users see messages in a specific order (sometimes known as surround sessions
- Excluding competitive creative so users do not see competitors' ads directly next to one another
- Displaying creatives so an advertiser can own 100% of the inventory on a page (sometimes known as roadblocks
- Targeting creatives to users based on their previous behavior (Behavioral marketing)

Pop-up ad

Pop-up ads are a form of online advertising on the World Wide Web intended to increase web traffic or capture email addresses. It works when certain web sites open a new web browser window to display advertisements. The pop-up window containing an advertisement is usually generated by JavaScript, but can be generated by other means as well.

A variation on the pop-up window is the **pop-under** advertisement. This opens a new browser window, behind the active window. Pop-unders interrupt the user less, but are not seen until the desired windows are closed, making it more difficult for the user to determine which Web site opened them.

Background

For early advertising-supported web sites, banner ads were sufficient revenue generators, but in the wake of the dot com crash, prices paid for banner advertising clickthroughs decreased and many vendors began to investigate more effective advertising methods. Pop-up ads by their nature are difficult to ignore or overlook, and are claimed to be more effective than static banner ads. Pop-ups have a much higher click rate than web banner ads do (about every 14,000th popup ad is clicked on).

Pornographic web sites are among the most common users of pop-up ads. Some particularly vicious types of pop-up ads (again, most often seen in connection with adult entertainment sites) appear to have either been programmed improperly or have been specifically designed to "hijack" a user's Internet session. These forms of pop-ups sometimes spawn multiple windows, and as each window is closed by the user it activates code that spawns another window -- sometimes indefinitely. This is sometimes referred to by users as a "Java trap", "spam cascade" or "Pop-up Hell" among other names. Usually the only way to stop this is to close the browser.

Another variation of pop-up, commonly called "mousetrapping", particularly fills an entire screen with an ad or Web page, in the process removing any menu bars or other on-screen icons by which the user can close the window. This problem mainly affects users of the Windows version of Internet Explorer. Often, access to other open windows and Web pages is denied. One way for PC users to close these ad windows is via the control-alt-delete command, which can result in all active IE windows (including those not connected to the pop-up) closing, and another way to close the mousetrapping window could be to hold down the Alt button and press F4 to close the active window. Another variant, a "static image ad", is a pop-up ad that stays in a fixed position of a window of an ad-supported program. This kind of ad does not distract the computer's concentration of a program window like a traditional popup ad does. One example of an ad-supported program that uses a static image ad is KaZaA.

Non-browser pop-up ads

Processes other than the Web browser can also display pop-up ads, or can direct the browser to display them. Many spyware programs do this, as well as some advertising-supported software, although the line between the two is sometimes thin.

A different sort of pop-up ad can be sent via the Messenger service in Microsoft's Windows operating system. These pop-ups appear as Windows dialog boxes with a textual message inside, usually directing the user to a Web site. Claims have been made that this type of pop-up has been used to commit extortion. Threats of legal action against the company D Squared Solutions has caused them to stop using this technique.

Popup generators

Popup generator is a term most often used to describe a web advertising software application. Popup generators are programs that create or generate Internet web-based advertisements also called popup ads. There are many types of popup ads, but there are two major categories:

1. Web browser popup ads based on a new browser window.
2. Web browser popup ads based on a JavaScript generated DHTML code that creates a pseudo-window inside the browser window.

Background

In the early stage of popup ads on the Internet, it was noticed by advertisers that popup advertisements draw more attention from visitors than any other type of advertisement would do, like banner ads, movie commercials, animated GIF images, audio commercials, textual ads, etc. In this same early stage, popup ads obtained a bad name because of hacker programmers who exploited browser bugs and holes. These programmers used to illegally install on the users' computers small pieces of software that would generate popup ads without using the Internet or the web browser. These were atypical popup windows that where not of the two standard sorts as defined in the beginning. Happily, with the time passed, browser vendors managed to fix the security holes and made their browsers stable enough not to let abuse practices for illegal advertising. Nowadays, almost all popup ads seen are legal, which means they are part of the web content and are generated with legal and documented web technologies.

Are Popup Ads Annoying?

Certainly, sometimes popup ads are annoying, just as any other advertisement may be. But classifying popup ads as annoying in general would be rather subjective assumption, close to classifying TV commercials as being annoying, because they interrupt the movies in the most interesting part. There are people that love commercials, especially if the latter are aesthetic and created with artistic taste.

Recent works

Being still one of the most efficient web advertising technologies, popup ads and popup generator applications experience constant development and improvement. Some of the latest developers find ways of making popup generators that create pleasant popup ads instead of annoying ones. These generators create the so-called Hover Ads. The major technology used to produce these ads is JavaScript along with other new web techniques. These popups represent normal web browser content also called DHTML content and avoid opening new browser windows. Modern popup generators utilize likable visual effects, which tends to stabilize the effectiveness of this sort of Internet advertising.

Hover Ads

Hover Ads are a special type of popup ads created using JavaScript and similar web browser technologies.

Background

Being the most effective way of web advertising, popup ads acquired a prominent share of web advertising solutions and technologies. The first Internet popups were created using the window.open() JavaScript function, which opens a new browser window. These popups were easily blocked by popup blockers and web advertising specialists started looking for new approaches. As such were used Macromedia/Flash ads and Hover ads popup generators.

Technology

The latter are developed around several web browser technologies but in the center of their realization is utilized the so called modern form of HTML – DHTML. DHTML is a synthesis between HTML language and JavaScript language. Using JavaScript, certain levels and objects of the browser's DOM are manipulated to produce window-like visual DHTML elements representing Hover ads or hover ad windows.

Web banner

A **web banner** or **banner ad** is a form of advertising on the World Wide Web. This form of online advertising entails embedding an advertisement into a web page. It is intended to attract traffic to a website by linking them to the web site of the advertiser. The advertisement is constructed from an image (GIF, JPEG), JavaScript program or multimedia object employing technologies such as Java, Shockwave or Flash, often employing animation or sound to maximize presence. Images are usually in a high-aspect ratio shape. That is to say, either wide and short, or tall and narrow, hence the reference to banners. These images are usually placed on web pages that have interesting content, such as a newspaper article or an opinion piece.

The web banner is displayed when a web page that references the banner is loaded into a web browser. This event is known as an "impression". When the viewer clicks on the banner, the viewer is directed to the website advertised in the banner. This event is known as a "click through". In many cases, banners are delivered by a central ad server.

Many banner ads work on a click-through payback system. When the advertiser scans their logfiles and detects that a web user has visited the advertiser's site from the content site by clicking on the banner ad, the advertiser sends the content provider some small amount of money (usually around five to ten US cents). This payback system is often how the content provider is able to pay for the internet access to supply the content in the first place.

Web banners function the same way as traditional advertisements are intended to function: notifying consumers of the product or service and presenting reasons why the consumer should choose the product in question, although web banners differ in that the results for advertisement campaigns may be monitored real-time and may be targeted to the viewer's interests.

Many web surfers regard these advertisements as highly annoying because they distract from a web page's actual content or waste bandwidth. Newer web browsers often include

options to disable pop-ups or block images from selected websites. Another way of avoiding banners is to use a proxy server that blocks them, such as Privoxy.

History

The first clickable web ad (which later came to be known by the term "banner ad") was sold by Global Net Navigator (GNN) in 1993.

Founded by O'Reilly and Associates, Global Network Navigator (GNN) was the first commercially supported web publication and one of the very first web sites ever. Dale Dougherty was GNN's developer and publisher. O'Reilly and Associates sold GNN to AOL in 1995 and the site was discontinued a few years later.

The first web banner sold by HotWired, an important early pioneer in commercial web publishing started by Wired Magazine, was paid for by AT&T, and was put online on October 25, 1994.

HotWired was the first web site to sell banner ads in large quanities to a wide range of major corporate advertisers. Andrew Anker was HotWired's first CEO. Rick Boyce, a former media buyer with San Francisco advertising agency Hal Riney & Parnters, spearheaded the sales effort for the company. When HotWired was sold to Lycos, Boyce became its Vice President of Sales.

HotWired coined the term "banner ad" and was the first company to provide click through rate reports to its customers.

Standard sizes

The Interactive Advertising Bureau has released a set of sizes which it has designed to make ad sizing more predictable and better for both consumer and producer. It calls these web advertisements "interactive marketing units". The sizes are as follows (measurements in pixels):

- Sizes for rectangular/pop-up ads
    - Medium Rectangle: 300 by 250
    - Square Pop-Up: 250 square
    - Vertical Rectangle: 240 by 400
    - Large Rectangle: 336 by 280
    - Rectangle: 180 by 150
- Sizes for banner/button ads
    - Full Banner: 468 by 60
    - Half Banner: 234 by 60
    - Micro Button: 80 by 15
    - Micro Bar: 88 by 31
    - Button 1: 120 by 90
    - Button 2: 120 by 60
    - Vertical Banner: 120 by 240
    - Square Button: 125 square
    - Leaderboard: 728 by 90
- Sizes for "skyscraper" ads

- Wide Skyscraper: 160 by 600
- Skyscraper: 120 by 600
- Half Page Ad: 300 by 600

The IAB has also further standardized four of the sizes (Medium Rectangle, Rectangle, Leaderboard, Wide Skyscraper) into a set of guidelines it calls the "Universal Ad Package".

Types of web banners

Message Plus Unit (MPU)

A **Message Plus Unit** takes the form of a square advert that usually occurs in the middle of ordinary page content.

In the UK, "The Guardian" newspaper claim to be the first site to employ technology to automatically hide MPU slots that are not in use.

Ad filtering

Ad filtering or **ad blocking** is a service which removes or alters advertising content in a webpage. This content can be represented in a variety of ways including pictures, animations, text, or pop-up windows. More advanced filters allow fine-grained control of advertisements through features like blacklists, whitelists, and regular expression filters. Certain security features also have the effect of disabling some ads.

The immediate benefits include cleaner looking webpages free from advertisements and lower resource-usage (bandwidth, CPU, memory, etc.). One drawback is that advertisements are a major source of revenue for many websites. However, the actual loss of revenue, when present, is difficult to measure.

Browser integration

Some web browsers support ad filtering through built-in features and plugins. A number of popular browsers include a pop-up blocker, such as Microsoft's Internet Explorer, Mozilla Firefox , Opera Software's Opera, and Apple Computer's Safari. All of these browsers support extensions and/or plugins which can include ad filters. For example, Adblock is a popular extension for Firefox.

External programs

A number of external applications offer ad filtering as a primary or additional feature. A traditional solution is to customize an HTTP proxy (or web proxy) to filter content. Proxies may reside on and serve a single computer or serve a number of clients over a network. These programs work by caching and filtering content before it is displayed in a user's browser. This provides an opportunity to remove not only ads, but content which may be offensive or inappropriate. Popular proxy software which can be used as effective ad filters include: Privoxy, Squid, Proximodo, and Proxomitron.

Common advertising techniques

- Pop-up ads
- Plain text
- Ad banners
- Flash animations
- Keyword hyperlinks (for example Vibrant Media's IntelliTXT)
- Browser plugins/extensions (often labeled as adware)
- External applications (see adware, spyware).

Pop-up blocking

Opera was the first major browser to incorporate popup-blocking tools; the Mozilla browser later improved on this by blocking only popups generated as the page loads. In the early 2000s, all major web browsers except Internet Explorer (then the most popular browser and still as of 2006) allowed the user to block unwanted pop-ups almost completely. In 2004, Microsoft released Windows XP SP2, which added pop-up blocking to Internet Explorer. Many users, however, remain unaware of this ability, or else choose not to use it. Many others are not able to use it at all, as they do not use Windows XP SP2, but older versions of Windows. Some users install non-Microsoft ad-blocking software instead.

Most modern browsers come with pop-up blocking tools; third-party tools tend to include other features such as ad filtering.

Problems with pop-up blockers and non-advertising 'pop-ups'

Cyworld is one of the largest Korean communities on the web, with approximately 11 million users. Each user has a home page, pre-designed and the same size, but customizable. The home page itself, however, is technically a pop-up as it is less than the size of a typical browser window (a so-called *mini hompy*, or miniature home page). After Windows XP SP2 was released, there was a flurry of activity as Cyworld changed its front page to explain to its 11 million users (nearly a quarter of the population) how to get past the pop-up blocker.

Circumventing pop-up blockers

Advertisers continually seek ways to circumvent such restrictions. Many of the latest pop-ups are created using Flash and have extensive animation and trickery; others use DHTML to appear in front of the browser screen.

A form of advertisement that combines elements of a pop- up and web banner is a Flash animation superimposed over a webpage in a transparent layer. The flash animation links to the advertiser's site or product. This is a new form of advertisement, created in response to the growing popularity of pop-up blockers. Because the advertisement is an embedded flash object, it can be blocked, but with more difficulty, as most programs would view it as part of the content of the page. Methods of removing these are by using CSS, or third-party extensions such as Adblock.

On the other hand, the so called Hover Ads or DHTML pop-ups are based primarily on the JavaScript browser capabilities. Certain popup generators utilize JavaScript code that creates DOM object elements organized in a system that uses CSS and mostly the position attribute, but not solely, to produce emulative behaviors and visual effects resembling windows with chrome, content and other attributes and effects unavailable for the old fashioned and blockable popups. This technology and approach of creating popups seems to be hardest to block, as it is a fluent part of the browser's HTML and DHTML content. A way of blocking hover ads is by disabling JavaScript of the browser, but this action leads to crippling the browser and is unacceptable.

Payment

Cost Per Impression

Cost Per Impression is a phrase often used in online advertising and marketing related to web traffic. It is used for measuring the worth and cost of a specific e-marketing campaign. This technique is applied with web banners, text links, e-mail spam, and opt-in e-mail advertising. (Although opt-in e-mail advertising is more commonly charged on a CPA basis.)

The *Cost Per Impression* is often measured using the **CPM** (Cost Per Mille) metric. (A *CPM* is the cost of one thousand (1,000) impressions.)

CPM is considered the optimal form of selling online advertising from the publisher's point of view. A publisher gets paid for each ad that is shown.

This type of advertising arrangement closely resembles Television and Print Advertising Methods for speculating the cost of an Advertisement. With Television the Nielsen Ratings are used and Print is based on how many readers a publication has. For a Website the numbers are a bit more exact due to the TCP/IP nature of the Internet.

CPM and/or Flat rate advertising deals are preferred by the Publisher/Webmaster because they will get paid regardless of any action taken.

For Advertisers a Performance Based system is preferred. There are two methods for Paying for Performance: 1) CPA - Cost per Action/Acquisition and 2) CPC - Cost per Click Through.

Today, it is very common for large publishers to charge for most of their advertising inventory on a *CPM* or CPT basis.

A related term, eCPM or effective Cost Per Mille, is used to measure the effectiveness of advertising inventory sold (by the publisher) via a CPC, CPA, or CPT basis.

Cost Per Mille

The initialization *CPM* comes from print world (and is a latin word), and stands for *Cost Per Mille* in the US or, more correctly, in the UK *Cost Per M*, with *M* representing the Roman numeral for thousand. When online advertising started gaining momentum, those in the industry

used this term (rather than something like *CPI*) as a metric for describing the *Cost Per Impression* largely because advertisers were already familiar with the term *CPM*.

It is important to remember that when someone says something like, "our CPM is $5". That this means that the *Cost Per Impressions* is $0.005 -- half a cent.

Cost Per Thousand

Cost per Thousand (known as CPM) is used in marketing as a benchmark to calculate the relative cost of an advertising campaign or an ad message in a given medium. Rather than an absolute cost, CPM estimates the cost per 1000 views of the ad.

It is calculated by:

total cost * 1000 / total audience

For example, while the Super Bowl has the highest per-spot ad cost in the United States, it also has the most television viewers annually. Consequently, its CPM may be comparable to a less expensive spot aired during standard programming.

The "M" in CPM derives from the Latin *mille* for "thousand."

In the United Kingdom, Cost Per Thousand is expressed as CPT rather than CPM.

Effective Cost Per Mille

Effective Cost Per Mille or **eCPM** (as it is often initialized to) is a phrase often used in online advertising and online marketing circles. It means the cost of every 1,000 ad impressions shown.

CPM is considered the optimal form of selling online advertising from the publisher's point of view. A publisher gets paid every time an ad is shown.

eCPM is used to measure the effectiveness of a publisher's inventory being sold (by the publisher) via a CPA, CPC, or CPT basis. In other words, the **eCPM** tells the publisher what they would have received if they sold the advertising inventory on a *CPM* basis (instead of a CPA, CPC, or CPT basis).

Cost Per Action

Cost Per Action or **CPA** (as it is often initialized to) is a phrase often used in online advertising and online marketing circles.

CPA is considered the optimal form of buying online advertising from the advertiser's point of view. An advertiser only pays for the ad when an *action* has occurred. An *action* can be a product being purchased, a form being filled, etc. (The desired *action* to be preformed is determined by the advertiser.)

A related term, eCPA or effective Cost Per Action, is used to measure the effectiveness of advertising inventory purchased (by the advertiser) via a CPC, CPM, or CPT basis.

The *CPA* can be determined by different factors, depending where the online advertising inventory is being purchased.

Other common forms, of charging for advertising, include:

- CPC
- CPM
- CPT

Effective Cost Per Action

Effective Cost Per Action (often abbreviated to **eCPA**) is a phrase often used in online advertising and online marketing circles.

CPA is considered the optimal form of buying online advertising from the advertiser's point of view, as they only pay for an advert when an *action* has occurred. An *action* can be a product being purchased, a form being filled, etc. (The desired *action* to be performed is determined by the advertiser.)

eCPA is used to measure the effectiveness of advertising inventory purchased (by the advertiser) via a CPC, CPM, or CPT basis. In other words, the **eCPA** tells the advertiser what they would have paid if they purchased the advertising inventory on a *CPA* basis (instead of a CPC, CPM, or CPT basis).

Cost Per Click

Cost Per Click or **CPC** (as it is often initialized to) is a phrase often used in online advertising and online marketing circles.

With many advertising networks and websites, the advertiser is charged for advertising their ad (on the advertising network or website) only when a user clicks on their ad. How much they pay (for that click) is called their *Cost Per Click* or *CPC*.

The *CPC* can be determined by different factors, depending on which advertising network or website the advertiser is advertising on.

Other common forms, of charging for advertising, include:

- CPM
- CPA
- CPT

Pay per click

Pay per click, or PPC, is an advertising technique used on websites, advertising networks, and search engines.

With search engines, pay per click advertisements are usually text ads placed near search results; when a site visitor clicks on the advertisement, the advertiser is charged a small amount. Variants include pay for placement and pay for ranking. Pay per click is also sometimes known as Cost Per Click (CPC).

While many companies exist in this space, Google Adwords and Yahoo! Search Marketing, which was formerly Overture, are the largest network operators as of 2006. MSN has

started beta testing with their own PPC services MSN adCenter. Depending on the search engine, minimum prices per click start at US$0.01 (up to US$0.50). Very popular search terms can cost much more on popular engines. Abuse of the pay per click model can result in click fraud. Click fraud is usually not detected very well by smaller PPC engines.

Categories

PPC engines can be categorized in "Keyword", "Product", "Service" engines. However, a number of companies may fall in two or more categories. More models are continually being developed.

Keyword PPCs
Advertisers using these bid on "keywords", which can be words or phrases, and can include product model numbers. When a user searches for a particular word or phrase, the list of advertiser links appears in order of bidding.

As of 2005, notable PPC Keyword search engines include: Google AdWords, Yahoo! Search Marketing, GaZabo.com, Miva, which was formerly FindWhat, SearchFeed, Enhance (formerly Ah-Ha), GoClick, 7Search, Kanoodle, ePilot, Search123, Kazazz, Pricethat, Search FAST and others.

An industry of professional services firms that can assist advertisers in marketing their products and services on search engines has also developed. Many of these firms will be members of various trade bodies such as IABUK, SMA-UK and SEMPO, while other reputable firms have chosen to avoid these bodies, as many of them remain heavily biased toward the firms that first got together and founded them.

Product PPCs
"Product" engines let advertisers provide "feeds" of their product databases and when users search for a product, the links to the different advertisers for that particular product appear, giving more prominence to advertisers who pay more, but letting the user sort by price to see the lowest priced product and then click on it to buy. These engines are also called Product comparison engines or Price comparison engines.

Some of the PPC Product search engines are: BizRate, NexTag, PriceGrabber, Pricescan, Pricethat, Pricewatch, PriceLeap, Shopping.com

Service PPCs
"Service" engines let advertisers provide feeds of their service databases and when users search for a service offering links to advertisers for that particular service appear, giving prominence to advertisers who pay more, but letting users sort their results by price or other methods. Some Product PPCs have expanded into the service space while other service engines operate in specific verticals.

Examples of PPC services include NexTag, Pricethat SideStep, and TripAdvisor.

Pay per Call
Similar to pay per click, pay per call is a business model for ad listings in search engines and directories that allows publishers to charge local advertisers on a per-call basis for each lead (call) they generate. The term "pay per call" is sometimes confused with "click to

call". Click-to-call, along with call tracking, is a technology that enables the “pay-per-call” business model.

Click-through rate

Click-through rate or **CTR** is a way of measuring the success of an online advertising campaign. A CTR is obtained by dividing the number of users who clicked on an ad on a web page by the number of times the ad was delivered (impressions). For example, if your banner ad was delivered 100 times (impressions delivered) and 1 person clicked on it (clicks recorded), then the resulting CTR would be 1%.

Banner ad click-through rates have fallen over time, often measuring significantly less than 1%. By selecting an appropriate advertising site with high affinity (e.g. a movie magazine for a movie advertisement), the same banner can achieve a substantially higher click-through rate. Personalized ads, unusual formats, and more obtrusive ads typically have higher click-through rates than standard banner ads.

Click fraud

Click fraud occurs in pay per click online advertising when a person, automated script or computer program imitates a legitimate user of a web browser clicking on an ad, for the purpose of generating an improper charge per click. Click fraud is the subject of some controversy and increasing litigation due to the advertising networks being a key beneficiary of the fraud whether they like it or not.

Use of a computer to commit this type of fraud is a felony in many jurisdictions, for example as covered by Penal code 502 in California and the Computer Misuse Act 1990 in the United Kingdom. There have been arrests relating to click fraud with regard to malicious clicking in order to deplete a competitor's advertising budget.

In 2004, a California man created a software program that he claimed could let spammers defraud Google out of millions of dollars in fraudulent clicks. Authorities said he was arrested while trying to blackmail Google for $150,000 to hand over the program.

Pay per click advertising

Pay per click advertising or **PPC advertising** is when webmasters (operators of web sites), acting as **publishers**, display clickable links from **advertisers**, in exchange for a charge per click. As this industry evolved, a number of **advertising networks** developed, who acted as middlemen between these two groups (publishers and advertisers). Each time a (believed to be) valid web user clicks on an ad, the advertiser pays the advertising network, who in turn pays the publisher a share of this money. This revenue sharing system is seen as an incentive for click fraud.

The largest of the advertising networks, Google's AdWords/AdSense and Yahoo! Search Marketing, act in a dual role, since they are also publishers themselves (on their search engines). According to critics, this complex relationship may create a conflict of interest.

For instance, Google loses money to undetected click fraud when it pays out to the publisher, but it makes money, when it collects it from the advertiser.

Non-contracting parties

A secondary source of click fraud is **non-contracting parties**, who are not part of any pay-per-click agreement. This type of fraud is even harder to police because perpetrators generally can not be sued for breach of contract, or charged criminally with fraud. Examples of non-contracting parties are:

- **Competitors of advertisers**: These parties may wish to harm a competitor who advertises in the same market by clicking on their ads. The perpetrators don't profit directly, but force advertiser to pay for irrelevant clicks, thus weakening or eliminating a source of competition.
- **Competitors of publishers**: These persons may wish to frame a publisher. It is made to look like the publisher is clicking on their own ads. The advertising network may then terminate the relationship. Many publishers rely exclusively on revenue from advertising, and can be put out of business by such an attack.
- **Other malicious intent**: As with vandalism, there's an array of motives for wishing to cause harm to either an advertiser or a publisher, even by people who have nothing to gain financially. Motives include political and personal vendettas. These cases are often the hardest to deal with, since it is hard to track down the culprit, and if found, there is little legal action that can be taken against them.
- **Unwanted "friends" of the publisher**: Sometimes upon learning a publisher profits from ads being clicked, a supporter of the publisher (like a fan, family member, or personal friend), will click on the ads, to "help". However, this can backfire when the publisher (not the "friend") is accused of click fraud.

Advertising networks try to stop fraud by all parties, but often do not know which clicks are legitimate. Unlike fraud committed by the publisher, it is hard to know who should pay when past click fraud is found. Publishers resent having to pay refunds for something that is not their fault. However, advertisers are adamant that they should not have to pay for phony clicks.

Organization

Click fraud can be as simple as one person starting a small web site, becoming a publisher of ads, and clicking on those ads to generate revenue. Oftentimes, the number of clicks, and their value, is so small, that the fraud goes undetected. Oftentimes publishers will claim small amounts of such clicking is an accident, which is often the case.

Much larger scale fraud also occurs. Those engaged in large scale fraud will often run scripts, which simulate a human clicking on ads in web pages. However, huge numbers of clicks appearing to come from just one, or a small number, of computers, or single geographic area, look highly suspicious to the advertising network and advertisers. Clicks coming from a computer known to be that of a publisher, also look suspicious to those

watching for click fraud. A person attempting large scale fraud, alone in their home, stands a good chance of being caught.

Organized crime can handle this by having many computers, with their own internet connection, in different geographic locations. Often scripts fail to mimic true human behavior, so organized crime networks use Trojan code to turn the average person's machines into zombie computers and using sporadic redirects or DNS-cache-poisoning to turn the oblivious user's actions into actions generating revenue for the scammer.

Impression fraud is an insidious variant of click fraud where the advertiser is penalized for having an unacceptably low click-through rate for a given keyword. This involves making numerous searches for a keyword but without clicking of the ad. Such keywords are disabled automatically, enabling a competitor's lower-bid ad for the same keyword to continue while several high bidders (on the first page of the search results) have been eliminated.

It is very difficult for advertisers, advertising networks, and authorities to pursue cases against networks of people spread around multiple countries.

Litigation

Disputes over the issue have resulted in a number of lawsuits. In one case, Google (acting as both an advertiser and advertising network) won a lawsuit against a Texas company called Auction Experts (acting as a publisher), which Google accused of paying people to click on ads that appeared on Auction Experts' site, costing advertisers $50,000. Despite networks' efforts to stop it, publishers are suspicious of the motives of the advertising networks, because the advertising network receives money for each click, even if it is fraudulent.

Solutions

Proving click fraud can be very difficult, since it is hard to know who is behind a computer and what their intentions are. Often, the best an advertising network can do is to identify which clicks are most likely fraudulent, and not charge the account of the advertiser. Ever more sophisticated means of detection are used, but none are foolproof.

The pay-per-click industry is lobbying for tighter laws on the issue. Many hope to have laws that will cover those not bound by contracts.

A number of companies are developing viable solutions for click fraud identification and are developing intermediary relationships with advertising networks. Such solutions fall into two categories:

a) Forensic analysis of advertisers' web server log files

This analysis of the advertiser's web server data requires an in-depth look at the source and behavior of the traffic. As industry standard log files are used for the analysis, the data is verifiable by advertising networks.

b) Third-party corroboration

Third parties offer web-based solutions that might involve placement of single-pixel images or Javascript on the advertiser's web pages and suitable tagging of the ads. The visitor may

be presented with a cookie. Visitor information is then collected in a third-party data store and made available for download. The better offerings make it easy to highlight suspicious clicks and they show the reasons for such a conclusion. Since an advertiser's log files can be tampered with, their accompaniment with corroborating data from a third party forms a more convincing body of evidence to present to the advertising network.

Spam

Spamming is commonly defined as the sending of unsolicited bulk e-mail - that is, email that was not asked for (unsolicited) by multiple recipients (bulk). A further common definiton of spam restricts it to unsolicited *commercial* e-mail, a definiton that does not consider non-commercial solicitations such as political or religious pitches, even if unsolicited, as spam.

In the popular eye, the most common form of spam is that delivered in e- mail as a form of commercial advertising. However, over the short history of electronic media, people have spammed for many purposes other than the commercial, and in many media other than e-mail. Spammers have developed a variety of spamming techniques, which vary by media: e-mail spam, instant messaging spam, Usenet newsgroup spam, Web search engine spam, spam in blogs, and mobile phone messaging spam.

Spamming is economically viable because advertisers have effectively no operating costs beyond the management of their mailing lists. Because the barrier to entry is so low, the volume of unsolicited mail has produced other costs which are borne by the public (in terms of lost productivity and fraud) and by Internet service providers, which must add extra capacity to cope with the deluge. Spamming is widely reviled, and has been the subject of legislation in a number of jurisdictions.

Solutions to the spam problem

All manner of attempts have been made to curb unsolicited mass electronic communications. For more information, see Stopping e-mail abuse. There are many solution categories in this constantly evolving field. Source-based blocking solutions prevent receipt of spam, while content filtering solutions identify spam after it's been received. There are avoidance strategies, including disposable identities. Automated cancellation of netnews spam is ongoing. Contractual measures such as Internet Service Providers' acceptable-use policies are also employed. Anti-spam laws such as the CAN-SPAM Act of 2003 have also been introduced to regulate or increase the legal penalties for spamming. Various vigilante and retaliatory tactics are also employed. Newer strategies include various cost-based and e-mail authentication and sender reputation solutions. The best means however is to be vigilant as to whom you give your email address. Constant distribution of your email address is bound to result in spam in some way. The best frame of mind is to decide whether the website can be trusted with your email address.

Spamming in different media

E-mail spam

E-mail spam is by far the most common form of spamming on the internet. It involves sending identical or nearly identical unsolicited messages to a large number of recipients. Unlike legitimate commercial e-mail, spam is generally sent without the explicit permission of the recipients, and frequently contains various tricks to bypass e-mail filters. Modern

computers generally come with some ability to send spam. The only necessary added ingredient is the list of addresses to target.

Spammers obtain e-mail addresses by a number of means: *harvesting* addresses from Usenet postings, DNS listings, or Web pages; guessing common names at known domains (known as a *dictionary attack*); and *"e-pending"* or searching for e-mail addresses corresponding to specific persons, such as residents in an area. Many spammers utilize programs called web spiders to find e-mail addresses on web pages, although it is possible to fool the web spider by substituting the "@" symbol with another symbol, for example "#", while posting an e-mail address.

Many e-mail spammers go to great lengths to conceal the origin of their messages. They might do this by spoofing e-mail addresses (similar to Internet protocol spoofing). In this technique, the spammer modifies the e-mail message so it looks like it is coming from another e-mail address. However, many spammers also make it easy for recipients to identify their messages as spam by placing an ad phrase in the *From* field—very few people have names like "GetMyCigs" or "Giving away playstation3s"!

Among the tricks used by spammers to try to circumvent the filters is to intentionally misspell common spam filter trigger words. For example, "viagra" might become "vaigra", or other symbols may be inserted into the word as in "v/i/a/g./r/a". The human mind can handle a surprising degree of corruption, but sometimes this tactic can backfire, rendering a message illegible. ISPs have begun to use the misspellings themselves as a filtering test.

The most dedicated spammers—often those making a great deal of money or engaged in illegal activities, such as the pornography, casinos and Nigerian scammers—are often one step ahead of the ISPs. Reporting them to your ISP may help block less sophisticated spammers in the future.

So-called "spambots" are a major producer of e-mail spam. The worst spammers create e-mail viruses that will render an unprotected PC a "zombie computer"; the zombie will inform a central unit of its existence, and the central unit will command the "zombie" to send a low volume of spam. This allows spammers to send high volumes of e-mail without being caught by their ISPs or being tracked down by antispammers; a low volume of spam is instead sent from many locations simultaneously. Many consumer-level ISPs (Earthlink, for example) stop spambots by blocking the SMTP port (port 25), although there are some users who make legitimate use of it.

Messaging spam

Messaging spam, sometimes termed *spim* (a portmanteau of spam and IM, short for instant messenger), makes use of instant messaging systems, such as AOL Instant Messenger or ICQ. Many IM systems offer a directory of users, including demographic information such as age and sex. Advertisers can gather this information, sign on to the system, and send unsolicited messages. To send instant messages to millions of users on most IM services merely requires scriptable software and the recipients' IM usernames. Spammers have similarly targeted Internet Relay Chat channels, using IRC bots that join channels and bombard them with advertising messages. Because most IM protocols are proprietary, it is easier to enact unilateral changes to make spamming more difficult.

A similar sort of spam can be sent with the Messenger Service in Microsoft Windows. The Messenger Service is an SMB facility intended to allow servers to send pop-up alerts to a Windows workstation. When Windows systems are connected to the Internet with this service running and without an adequate firewall, it can be used to send spam. The Messenger Service can, however, be easily disabled.

Messenger service spam, in particular, has lent itself to spammer use in a particularly circular scheme. In many cases, messenger spammers send messages to vulnerable Windows machines consisting of text like *"Annoyed by these messages? Visit this site."* The link leads to a Web site where, for a fee, users are told how to disable the Windows messenger service. Though the messenger service is easily disabled for free by the user, this scam works because it creates a perceived need and then offers an immediate solution. Oftentimes, the only "annoying messages" the user is receiving through Messenger are advertisements to disable Messenger itself.

Newsgroup spam and Forum spam

Newsgroup spam predates e-mail spam, and targets Usenet newsgroups. Old Usenet convention defines spamming as excessive multiple posting, that is, the repeated posting of a message (or substantially similar messages). Since posting to newsgroups is nearly as easy as sending e-mails, newsgroups are a popular target of spammers. The Breidbart Index was developed to provide an objective measure of the "spamminess" of a multi-posted or cross-posted message on Usenet.

Spamming an internet forum in general, is when a user posts something which is off-topic or doesn't have anything to do with the current subject. Also, a post that doesn't contribute to the thread whatsoever is also considered spam in some cases. A third form of Forum Spamming is where a person repeatedly posts about a certain subject in a manner that is unwanted by (and possibly annoying to) the general population of the forum. Lastly there is also the case where a person posts messages solely for the purpose of increasing his or her ranking on the forum. In a broader sense, advertising on forums where it is not wanted is known as spamming and is generally seen as an annoyance.

Mobile phone spam

Mobile phone spam is directed at the text messaging service of a mobile phone. This can be especially irritating to consumers not only for the inconvenience but also because they sometimes have to pay to receive the text message.

Internet telephony spam

It has been predicted that voice over IP (VoIP) communications will be vulnerable to being spammed by prerecorded messages. Although there have been few reported incidents, some companies have already tried to sell defenses against it.

Online game messaging spam

Many online games allow players to contact each other via player-to-player messaging, or chatrooms or public discussion areas.

What qualifies as spam varies from game to game, but usually this term is applied to all forms of flooding the game with messages; in case of MUDs, the problem is usually the same as with other chat .

Many games have strict rules on what kind of communication is acceptable in the games. Frequently, the terms of service don't allow promotion of external websites except on very strict terms (for example, URLs may be allowed on player profiles, but not anywhere else), and promotion of websites in-game is usually very much frowned on in any case.

Spam targeting search engines (Spamdexing)

Spamdexing (a portmanteau of *spamming* and *indexing*) refers to the practice on the World Wide Web of deliberately modifying HTML pages to increase the chance of them being placed high on search engine relevancy lists. People who do this are called search engine spammers. In layman's terms, spamdexing is using unethical means known as "black hat seo techniques" to unfairly increase the rank of sites in search engines. When a website is optimized to be indexable by a search engine, without trying to deceive its web crawler, this is called search engine optimization. To be sure, there is much gray area between *white-hat* search engine optimization and *black-hat* spamdexing.

Blog, wiki, guestbook, and referrer spam
Google's PageRank system uses the number of links to a page as an index of its "importance". Ordinarily, very few pages will link to a spammer's commercial site, because it is of no interest to anyone else, and hence it will have a very low PageRank score. To counter this effect, spammers attempt to create links to their sites on other people's pages.

The most common targets for this kind of spam are weblogs, the spamming then being known as blog spam, or "blam" for short. In 2003, this type of spam took advantage of the open nature of comments in the blogging software Movable Type by repeatedly placing comments to various blog posts that provided nothing more than a link to the spammer's commercial web site.

Similar attacks are often performed against wikis and guestbooks, both of which accept user contributions; something that consistantly impresses and confounds critics of Wikipedia is its remarkable lack of spam, in spite of having nearly one million articles and over two million pages.

On January 18, 2005, Google proposed a `rel="nofollow"` attribute that could be placed on a link; doing so instructs most major search engines to ignore the link, rendering it useless to spammers. Software is then rewritten to add this attribute to any link embedded in a comment. As of April 2005, nofollow has seen expanding usage, but is not yet universal.

As well as comment forms, editable pages and guestbooks, some sites publish a list of the most common referrers to their site in order to show how readers have found it. These lists have also been exploited by spammers with so- called referer spam, in which the spammer makes repeated web site requests using a fake referer URL pointing to a spam-advertised site. That URL will later appear as a link on the site, boosting the PageRank of its target.

Commercial uses

The most common purpose for spamming is advertising. Goods commonly advertised in spam include pornography, unlicensed computer software, medical products such as Viagra, credit card accounts, and fad products. In part because of the bad reputation (and dubious legal status) which spamming carries, it is chiefly used to carry offers of an ill-reputed or legally questionable nature. Many of the products advertised in spam are fraudulent in nature, such as quack medications and get-rich-quick schemes. Spam is frequently used to advertise scams, such as diploma mills, advance fee fraud, pyramid schemes, stock pump-and-dump schemes, and phishing. It is also often used to advertise pornography without regard to the age of the recipient, or the legality of such material in the recipient's location.

One of the most common ad spams is the computer software program GAIN. Also known as Gator or Claria or Dashbar, this insidious program hides itself within the active programs running on your computer and will collect information on internet habits. Based on the websites you visit, it will then send you "relevant" advertising at random intervals. Unfortunately, this program is often attached and automatically installed with popular "free" software, such as many P2P filesharing clients. Even removing GAIN from your computer can sometimes prove difficult, as it leaves traces of itself even after uninstallation or removal by third party spyware programs.

Spam has different levels of acceptability in different countries. For example, in Russia spamming is commonly used by many mainstream legitimate businesses, such as travel agencies, printing shops, training centers, real estate agencies, seminar and conference organizers, and even self-employed electricians and garbage collection companies. In fact, the most prominent Russian spammer was American English Center, a language school in Moscow. That spamming sparked a powerful antispam movement by enraging the Deputy Minister of Communications Andrey Korotkov and provoking a wave of counterattacks on the spammer through non-Internet channels, including a massive telephone DDOS (Distributed Denial of Service) attack.

Comparison to postal "junk" mail

There are a number of differences between spam and junk mail:

- Unlike junk postal mail, the costs of spam paid for by the recipient's mail site commonly approach or even exceed those of the sender, in terms of bandwidth, CPU processing time, and storage space. Spammers frequently use free dial-up accounts, so their costs may be quite minimal indeed. Because of this offloading of costs onto the recipient, many consider spamming to be criminal conversion or theft.
- Junk mail can be said to subsidize the delivery of mail customers want to receive. For example, the United States Postal Service allows bulk mail senders to pay a lower rate than for first-class mail, because they are required to sort their mailings and apply bar codes, which makes their mail much cheaper to process. While some ISPs receive large fees from spammers, most do not—and most pay the costs of carrying or filtering unwanted spam.

- Another distinction is that the costs of sending junk mail provide incentives to be somewhat selective about recipients, whereas the spammer has much lower costs, and therefore much less incentive.
- Finally, bulk mail is by and large used by businesses that are traceable and can be held responsible for what they send. Spammers frequently operate on a fly-by-night basis, using the so-called "anarchy" of the Internet as a cover

Spamdexing

Spamdexing or **search engine spamming** is the practice of deliberately creating web pages which will be indexed by search engines in order to increase the chance of a website or page being placed close to the beginning of search engine results, or to influence the category to which the page is assigned. Many designers of web pages try to get a good ranking in search engines and design their pages accordingly. The word is a portmanteau of *spamming* and *indexing*.

Spamdexing refers exclusively to practices that are dishonest and mislead search and indexing programs to give a page a ranking it does not deserve. "White hat" techniques for making a website indexable by search engines, without misleading the indexation process, are known as search engine optimization (SEO). SEO techniques do not involve deceit.

Search engine spammers, on the contrary, are generally aware that the content that they promote is not very useful or relevant to the ordinary internet surfer. Search engines use a variety of algorithms to determine relevancy ranking. Some of these include determining whether the search term appears in the META keywords tag, others whether the search term appears in the body text of a web page. A variety of techniques are used to spamdex (see below). Many search engines check for instances of spamdexing and will remove suspect pages from their indexes.

The rise of spamdexing in the mid-1990s made the leading search engines of the time less useful, and the success of Google at both producing better search results and combating keyword spamming, through its reputation-based PageRank link analysis system, helped it become the dominant search site late in the decade, where it remains. While it has not been rendered useless by spamdexing, Google has not been immune to more sophisticated methods either. Google bombing is another form of web vandalism, which involves creating pages that directly affect the rank of other sites.

Common **spamdexing** techniques can be classified into two broad classes: *content spam* and *link spam*.

Content spam

These techniques involve altering the logical view that a search engine has over the page's contents. They all aim at variants of the vector space model for information retrieval on text collections.

- **Hidden or invisible text**

  - Disguising keywords and phrases by making them the same (or almost the same) color as the background, using a tiny font size or hiding them within the HTML code such as "no frame" sections, ALT attributes and "no script" sections. This is useful to make a page appear to be relevant for a web crawler in a way that makes it more likely to be found. Example: A promoter of a Ponzi scheme wants to attract web surfers to a site where he advertises his scam. He places hidden text appropriate for a fan page of a popular music group on his page, hoping that the page will be listed as a fan site and receive many visits from music lovers.
- **Keyword stuffing**
  - This involves the insertion of hidden, random text on a webpage to raise the keyword density or ratio of keywords to other words on the page. Older versions of indexing programs simply counted how often a keyword appeared, and used that to determine relevance levels. Most modern search engines have the ability to analyze a page for keyword stuffing and determine whether the frequency is above a "normal" level.
- **Meta tag stuffing**
  - Repeating keywords in the Meta tags, and using keywords that are unrelated to the site's content.
- **Gateway** or **doorway pages**
  - Creating low-quality web pages that contain very little content but are instead stuffed with very similar key words and phrases. They are designed to rank highly within the search results. A doorway page will generally have "click here to enter" in the middle of it.
- **Scraper sites**
  - Scraper sites are created using various programs such as Traffic Equalizer. These programs are designed to 'scrape' search engine results pages and create 'content' for a website. These types of websites are generally full of clickable ads.

Link spam

Link spam takes advantage of link-based ranking algorithms, such as Google's PageRank algorithm, which gives a higher ranking to a website the more other highly-ranked websites link to it. These techniques also aim at influencing other link-based ranking techniques such as the HITS algorithm.

- **Link farms**
  - Involves creating tightly-knit communities of pages referencing each other, also known humorously as *mutual admiration societies*
- **Hidden links**
  - Putting links where visitors will not see them in order to increase link popularity.
- **Sybil attack**
  - This is the forging of multiple identities for malicious intent, named after a personality disorder with the same name. A spammer may create multiple

web sites at different domain names that all link to each other, such as fake blogs known as spam blogs.

- **Spam in blogs**
  - This is the placing or solicitation of links randomly on other sites, placing a desired keyword into the hyperlinked text of the inbound link. Guest books, forums, blogs and any site that accepts visitors comments are particular targets and are often victims of drive by spamming where automated software creates nonsense posts with links that are usually irrelevant and unwanted.
- **Spam blogs**
  - A spam blog, on the contrary, is a fake blog created exclusively with the intent of spamming.
- **Page hijacking**
- **Referer log spamming**
  - When someone accesses a web page, i.e. the referee, by following a link from another web page, i.e. the referer, the referee is given the address of the referer by the person's internet browser. Some websites have a referer log which shows which pages link to that site. By having a robot randomly access many sites enough times, with a message or specific address given as the referer, that message or internet address then appears in the referer log of those sites that have referer logs. Since some search engines base the importance of sites by the number of different sites linking to them, referer-log spam may be used to increase the search engine rankings of the spammer's sites, by getting the referer logs of many sites to link to them.
- **Buying expired domains**
  - Some link spammers monitor DNS records for domains that will expire soon, then buy them when they expire and replace the pages with links to their pages.

Some of these techniques may be applied for creating a Google bomb, this is, to cooperate with other users to boost the ranking of a particular page for a particular query.

Other types of spamdexing

- Mirror websites
  - Hosting of multiple websites all with the same content but using different URLs. Some search engines give a higher rank to results where the keyword searched for appears in the URL.
- Page redirects
  - Taking the user to another page without his or her intervention, e.g. using META refresh tags, CGI scripts, Java, JavaScript, Server side redirects or server side techniques.
- **Cloaking** refers to any of several means to serve up a different page to the search-engine spider than will be seen by human users. It can be an attempt to mislead search engines regarding the content on a particular web site. It should be noted, however, that cloaking can also be used to ethically increase accessibility of a site to users with disabilities, or to provide human users with content that search engines

aren't able to process or parse. It is also used to deliver content based on a user's location; Google themselves use IP delivery, a form of cloaking, to deliver results.

A form of this is '*code swapping*, this is: optimizing a page for top ranking, then swapping another page in its place once a top ranking is achieved.

The following techniques are also widely acknowledged as being spam, or "black hat":

1. Doorway pages
2. Link farms
3. Googleating

Cloaking

Cloaking is a search engine optimization technique in which the content presented to the search engine spider is different from that presented to the users' browser; this is done by delivering content based on the IP addresses or the User-Agent HTTP header of whatever is requesting the page. The only legitimate uses for cloaking used to be for delivering content to users that search engines couldn't parse, like Macromedia Flash. However, cloaking is often used as a spamdexing technique, to try to trick search engines into giving the relevant site a higher ranking; it can also be used to trick search engine users into visiting a site based on the search engine description which site turns out to have substantially different - or even pornographic - content. For this reason some search engines threaten to ban sites using cloaking.

Cloaking is a form of the doorway page technique.

- A similar technique is also used on the Open Directory Project web directory. It differs in several ways from search engine cloaking:
- It is intended to fool human editors, rather than computer search engine spiders.

The decision to cloak or not is based upon the HTTP referrer, which tells the URL of the page on which a user clicked a link to get to the page. Some cloakers will give the fake page to anyone who comes from a web directory website, since directory editors will usually examine sites by clicking on links that appear on a directory webpage. Other cloakers give the fake page to everyone *except* those coming from a major search engine; this makes it harder to detect cloaking, while not costing them many visitors, since most people find websites by using a search engine.

In more recent times several well known and well respected sites have taken up cloaking to deliver personalised content to their regular customers. In fact, many of the top 1000 sites - including household names like Amazon (amazon.com) - actively cloak. None of these have been banned from search engines purely because of cloaking.

Increasingly, for a page without natural popularity due to compelling or rewarding content to rank well in the search engines, Webmasters must design pages solely for the search engines. This results in pages with too many keywords and other factors that might be search engine "friendly", but make the pages difficult for actual endusers to consume. As such, cloaking is an important technique to allow Webmasters to split their efforts and

separately target the search engine spiders and endusers. As with anything, this technique can be used responsibly, or less so.

Page hijacking

Page hijacking is a form of spamdexing (spamming the index of a search engine). It is the act of copying a random but popular webpage with the intent to feed a web crawlers the copied page. The intent is that the copied page appears in search engine results, and when the users click on it, the visitors are redirected to a different, often unrelated, website.

Page hijacking is a form of cloaking, and it is possible because web crawlers detect duplicates as they download web pages, and if two pages have the same content, they keep only one of the URLs.

Doorway page

Doorway pages are web pages that are created for spamdexing, this is, for spamming the index of a search engine by inserting results for particular phrases with the purpose of sending you to a different page. They are also known as landing pages, bridge pages, portal pages, zebra pages, jump pages, gateway pages, entry pages and by other names. Doorway pages that redirect visitors without their knowledge use some form of cloaking.

If you click through to a typical doorway page from a search engine result page, in most cases you will be redirected with a fast Meta refresh command to another page. Other forms of redirection include use of Javascript and server side redirection, either through the .htaccess file or from the server configuration file. Some doorway pages may be dynamic pages generated by scripting languages such as Perl and PHP.

Doorway pages are often easy to identify in that they have been designed primarily for search engines, not for human beings. Sometimes a doorway page is copied from another high ranking page, but this is likely to cause the search engine to detect the page as a duplicate and exclude it from the search engine listings.

Because many search engines give you a penalty for using the META refresh command, some doorway pages just trick you into clicking on a link to get you to the desired destination page.

More sophisticated doorway pages, called *Content Rich Doorways*, are designed to gain high placement in search results without utilizing redirection. They incorporate at least a minimum amount of design and navigation similar to the rest of the site to provide a more human-friendly and natural appearance. Visitors are offered standard links as calls to action.

Many sites now use content rich doorway pages for their pay-per-click campaign landing pages. These doorway pages may employ server side scripting to count click-throughs, visits, and other user actions to assist with marketing data collection.

Scraper site

A **scraper site** is a website that pulls all of its own information from other websites. In essence, no part of a scraper site is original. In the last few years, and due to the advent of the Google Adsense publishing plan, scraper sites have proliferated at an amazing rate. Wikipedia is frequently a source of material for scraper sites.

Many scrapers will pull snippets and text from websites that rank high for keywords they've targeted. This way they hope to rank highly in the SERPS (Search Engine Results pages). RSS feeds are vulnerable to Scrapers.

Some scraper sites consist of advertisements and paragraphs of words randomly selected from a dictionary. Often a visitor will click on an advertisement because it is the only comprehensible text on the page. Operators of these scraper sites gain financially from these clicks. Ad networks such as Google Adsense are constantly working to remove these sites from their programs.

Scrapers tend to be associated in the mind with link farms and are sometimes perceived as the same thing.

Spam blogs

Spam blogs, sometimes referred to by the neologism **splogs**, are weblog sites which the author uses only for promoting affiliated websites. The purpose is to increase the PageRank of the affiliated sites, get ad impressions from visitors, and/or use the blog as a link outlet to get new sites indexed. Content is often nonsense or text stolen from other websites with an unusually high number of links to sites associated with the splog creator which are often disreputable or otherwise useless websites.

There is frequent confusion between the terms "splog" and "spam in blogs". **Splogs are blogs where the articles are fake**, and are only created for spamming.

To **spam in blogs**, on the contrary, is to include random comments on the blogs of innocent bystanders, in which spammers take advantage of a site's ability to allow visitors to post comments that may include links. This is used often in conjunction with other spamming techniques including *Sping*.

History

The term splog was popularized around mid August 2005 when it was used publicly by Mark Cuban, but appears to have been used a few times before for describing spam blogs going back to at least 2003. It developed from multiple linkblogs that were trying to influence search indexes and others trying to Google bomb every word in the dictionary.

Problems

Splogs have become a major problem on free blog hosts such as Google's Blogspot service. Some estimate it may be as high as one in five blogs These fake blogs waste valuable disk

space and bandwidth as well as pollute search engine results, ruining blog search engines and damaging bloggers community networking (e.g. Blogspot's next blog link). Google's search engine uses PageRank, which is very vulnerable to link flooding, especially from highly weighted bloggers. One splog clearly states: "Google's run by people who can be bothered to post links on the internet." Splogs could become a detractor to people using, enjoying and finding value in the blogosphere. Splogs sometimes choose a name similar to a popular blog. That way, they can benefit from the occasional incoming link from careless bloggers, who think they are linking to the popular site.

Benefits

They are good at launching new websites, as Google caches blogs frequently. There's rumored to be a secret webring called the 'Google brain' that does ethical splogging to improve the ranking of random good websites.

RSS abuse

Full content RSS feeds are actually compounding the splog problem. RSS makes it easy to steal content from genuine blogs. Splog RSS feeds pollute RSS search engines. Splog RSS feeds are being reproduced and plastered all over the net.

Defense

Several splog reporting services have been created for good willed users to report splog with plans of offering these splog URLs to search engines so that they can be excluded from search results. Splog Reporter was the first service of this kind. Then came SplogSpot which actually maintains a large database of splogs and makes it available to the public via APIs, and A2B which blocks web server IP addresses that splog URLs resolve to. As well as automated attempts to find them. Blogger has implemented a system that can detect splogs and then force them to take a Captcha 'spell this word' test. Blogger has recently deleted thousands of splogs in September and even more in December.

Spam in blogs

Spam in blogs (also called simply **blog spam** or **comment spam**) is a form of spamdexing. It is done by automatically posting random comments, promoting commercial services, to blogs, wikis, guestbooks, or other publicly-accessible online discussion boards. Any web application that accepts and displays hyperlinks submitted by visitors may be a target.

Adding links that point to the spammer's web site increases the page rankings for the site in the search engine Google. An increased page rank means the spammer's commercial site would be listed ahead of other sites for certain Google searches, increasing the number of potential visitors and paying customers.

History

This type of spam originally appeared in internet guestbooks, where spammers repeatedly fill a guestbook with links to their own site and no relevant comment to increase search

engine rankings. If an actual comment is given it is often just "cool page", "nice website", or keywords of the spammed link.

In 2003, spammers began to take advantage of the open nature of comments in the blogging software like Movable Type by repeatedly placing comments to various blog posts that provided nothing more than a link to the spammer's commercial web site. Jay Allen created a free plugin, called MT-BlackList, for the Movable Type weblog tool (versions prior to 3.2) that attempted to alleviate this problem. Many current blog software now have methods of preventing or reducing the effect of blog spam.

Possible solutions

rel=nofollow

In early 2005 Google announced that hyperlinks marked with rel="nofollow" would not influence the link target's ranking in the search engine's index.

(`rel=nofollow` actually tells a search engine "Don't score this link" rather than "Don't follow this link." This differs from the meaning of `nofollow` as used within a robots meta tag, which **does** tell a search engine: "Do not follow any of the hyperlinks in the body of this document.")

Using `rel=nofollow` is a much easier solution that makes the improvised techniques above irrelevant. Most weblog software now marks reader-submitted links this way by default (with no option to disable it without code modification). A more sophisticated server software could spare the nofollow for links submitted by trusted users like those registered for a long time or on a whitelist or with a high karma. Some server software adds `rel=nofollow` to pages that have been recently edited but omits it from stable pages, under the theory that stable pages will have had offending links removed by human editors.

Some weblog authors object to the use of `rel=nofollow`, arguing, for example, that

- Link spammers will continue to spam everyone to reach the sites that do not use `rel=nofollow`
- Link spammers will continue to place links for clicking (by surfers), even if those links are ignored by search engines.
- Google is advocating the use of rel=nofollow in order to reduce the effect of heavy inter-blog linking on page ranking

In particular, in the Wikipedia after a discussion it was decided not to use nofollow and to use a spam blacklist instead. In this way, Wikipedia contributes to the scores of the pages it links to, and expects editors to link to relevant pages.

Turing tests

Various methods requiring humans to do spamming by hand have been attempted. A variety of captcha gateways have been implemented, in an effort to prevent bots from submitting entries. Drawbacks to this are the annoyance it poses for regular users, the lack of any alternative for visually impaired users, and the ability of some advanced bots to fool simple captchas most of the time.

Server-side redirects

Instead of displaying a direct hyperlink submitted by a visitor, a web application could display a link to a script on its own website that redirects to the correct URL. This will not prevent all spam since spammers do not always check for link redirection but has proven very effective. Redirecting links prevent Google from factoring the link in its PageRank algorithm for that site making the spam ineffective. An added benefit is that the redirection script can count how many people visit external URLs, although it will increase the load on the site.

This kind of redirection can also be done via the .htaccess file in Apache, thus saving the load of a script.

Another way of preventing PageRank leakage without using client-side JavaScript or .htaccess file is the public redirection service like a TinyURL or My-Own.Net. For example,

Link

where 'alias_of_target' is the alias of target address.

Client-side redirects

Another option is for the script to be client-side JavaScript. For example,

Link

would work as a link but not be picked up by Google. Moreover, the javascript could be more complicated to ensure that the link would never be picked up since it was encoded. For example,

Link

where 'hfksksgjlsll' is an encoded URL that is decoded by the javascript function redirectFunction which presumably is stored in the HEAD tag of the page. A downside of this is that visitors who have disabled Javascript in their browser would be unable to follow the links.

Distributed Approaches

This approach is very new to addressing link spam. One of the shortcomings of link spam filters is that most sites only receive one link from each domain which is running a spam campaign. If the spammer varies IP addresses, there is little to no distiguishable pattern left on the vandalized site. The pattern, however, is left across the thousands of sites that were hit quickly with the same links.

A distributed approach, like the free LinkSleeve, uses XML-RPC to communicate between the various server applications (such as blogs, guestbooks, forums, and wikis) and the filter server, in this case LinkSleeve. The posted data is stripped of urls and each url is checked against recently submitted urls across the web. If a threshold is exceeded, a "reject" response is returned, thus deleting the comment, message, or posting. Otherwise, an "accept" message is sent.

A more robust distributed approach is Akismet, which uses a similar approach to LinkSleeve but uses API keys to assign trust to nodes and also has wider distribution as a

result of being bundled with the 2.0 release of WordPress. They claim over 140,000 blogs contributing to their system. Akismet libraries have been implemented for Java, Python, Ruby, and PHP, but its adoption may be hindered by the requirement of an API key and its commercial use restrictions.

Application-specific anti-spam methods

Particularly popular software products such as Movable Type and MediaWiki have developed their own custom anti-spam measures, as spammers focus more attention on targeting those platforms. Whitelists and blacklists that prevent certain IPs from posting, or that prevent people from posting content that matches certain filters, are common defenses. More advanced access control lists require various forms of validation before users can contribute anything like linkspam.

The goal in every case is to allow good users to continue to add links to their comments, as that is considered by some to be a valuable aspect of any comments section.

RSS feed monitoring
Some wikis allow you to access an RSS feed of recent changes or comments. If you add that to your news reader and set up a smart search for common spam terms (usually viagra and other drug names) you can quickly identify and remove the offending spam.

Sping

Sping is short for 'ping spam', and are related to fraudulent pings from blogs using trackbacks, called **trackback spam**. Pings are messages sent from blog and publishing tools to a centralized network service (Ping Server) providing notification of newly published posts or content. Spings, or ping spam, are pings that are sent from spam blogs, or are sometimes multiple pings in a short interval from a legitimate source, often tens or hundreds per minute, due to misconfigured software, or a wish to make the content coming from the source appear fresh.

Spings, like spam blogs are increasingly problematic for the blogosphere. Estimates from Dave Winer's Weblogs and Matt Mullenweg's ping-o-matic service have put the sping rate -- the percentage of pings that are sent from spam blogs -- well above 50%. A study from Ebiquity Group, UMBC confirms that these numbers are around 75%.

The term was coined by David Sifry from Technorati in his February 2006 State of the Blogosphere report.

Spam mass

Spam mass is defined as "the measure of the impact of link spamming on a page's ranking." The concept was developed by Zolt´an Gy¨ongyi and Hector Garcia-Molina of Stanford University in association with Pavel Berkhin and Jan Pedersen of Yahoo!. This paper expands upon their proposed TrustRank methodology.

The researchers developed a *good core* and a *bad core* of selected Web documents from which they measured spam mass across a collection of documents. Two types of measurements, *absolute mass* and *relative mass*, are used to compare groups of documents. The higher the mass measurements, the more likely the documents are to be equivalent to spam.

Thresholds

A threshold value is used to identify groups of documents as spam. If their relative mass value exceeds the threshold, the documents are considered to be spam. A second threshold for the PageRank values of the selected documents is applied. Only high PageRank documents are labelled as spam.

The purpose of the methodology is to identify spam documents with artificially inflated PageRank values.

Made For AdSense

Made for AdSense [MFA] websites are typically made by web scraping content from other websites and then monetizing the site using Google AdSense. This is also a derogatory term used to refer to websites that have no redeeming value except to get web visitors to the web site for the sole purpose of clicking on Google AdSense advertisements.

The problem with Made for AdSense sites is they are considered sites that are spamming the search engines and diluting the search results by providing surfers with less than satisfactory search results. The scraped content is considered redunandant to that which would be shown by the search engine under normal circumstances had no MFA site been found in the listings.

These types of sites are being eliminated in various search engines and currently show up as supplemental results instead of being displayed in the initial search results.

Bookmark spam

Bookmark spam is a type of spam affecting social bookmarking, social software websites such as del.icio.us.

Mainly affecting the website in a Spamdexing attack, impacting on which links are popular or recent. They also trying to attack Google pagerank similar to spam blog attacks.

To stop the pagerank influencing problem, del.icio.us was forced to block the google search bot from indexing its links

"One user tried to abuse to push the google ranking of some business sites. So, the homepage doesn't show links of ********* anymore" -the editor of one website

Link spam is currently a problem in del.icio.us/popular as there is no way to trust users and fake users can easily be created thus putting spam links in the most popular URLs.

Referer spam

Referer spam is a kind of spamdexing (spamming aimed at search engines). The technique involves making repeated web site requests using a fake referer url pointing to a spam-advertised site. Sites that publicize their access logs, including referer statistics, will then also link to the spammer's site.

This benefits the spammer because of the free link, and also gives the spammer's site improved search engine link placement due to link-counting algorithms that search engines use.

Technical solutions

As with e-mail spam, web site operators who receive unwanted referer spam may respond using filtering and blocking. Some web sites receive so many referer spam hits that they amount to a denial of service attack on the server because there are not enough resources left on the server to handle legitimate traffic.

Noncommercial spam

E-mail and other forms of spamming have been used for purposes other than advertisements. Many early Usenet spams were religious or political in nature. Serdar Argic, for instance, spammed Usenet with historical revisionist screeds. A number of evangelists have spammed Usenet and e-mail media with preaching messages. A growing number of criminals are also using spam to perpetrate various sorts of fraud, and in some cases have used it to lure people to locations where they have been kidnapped, held for ransom and even murdered.

DoS spam

Spamming has also been used as a denial of service ("DoS") tactic, particularly on Usenet. By overwhelming the readers of a newsgroup with an inordinate number of nonsense messages, legitimate messages can be lost and computing resources are consumed. Since these messages are usually forged (that is, sent falsely under regular posters' names) this tactic has come to be known as sporgery (from spam + forgery). This tactic has for instance been used by partisans of the Church of Scientology against the alt.religion.scientology newsgroup (see Scientology vs. the Internet) and by spammers against news.admin.net-abuse.email, a forum for mail administrators to discuss spam problems. Applied to e-mail, this is termed mailbombing. The Usenet Meow Wars (circa 1996) were DoS attacks on various newsgroups aimed at specific posters, thus disrupting the newsgroups where they were active. The DoS attacks launched by Hipcrime, which continue today, are more specifically crafted as DoS attacks on entire newsgroups. The alt.sex newsgroups were rendered virtually uninhabitable by commercial porn site spammers, partially for advertising purposes and partially to destroy a perceived free competitor. (This spawned the creation of the moderated, unspammable soc.sexuality newsgroups.)

In a handful of cases, forged e-mail spam has been used as a tool of harassment. The spammer collects a list of addresses as usual, then sends a spam to them signed with the name of the person he wishes to harass. Some recipients, angry that they received spam and seeing an obvious "source", will respond angrily or pursue various sorts of revenge against the apparent spammer, the forgery victim. A widely known victim of this sort of harassment was Joe's CyberPost, which has lent its name to the offense: it is known as a joe job. Such joe jobs have been most often used against antispammers: in more recent examples, Steve Linford of Spamhaus Project and Timothy Walton, a California attorney, have been targeted. Sometimes victims (such as ROKSO-listed spammers) are subscribed to lists that don't practice verified opt-in, such as magazine subscriptions and e-mail newsletters, a practise known as subscriptionbombing.

Spammers have also abused resources set up for purposes of anonymous speech online, such as anonymous remailers. As a result, many of these resources have been shut down, denying their utility to legitimate users.

E-mail worms or viruses may be spammed to set up an initial pool of infected machines, which then resend the virus to other machines in a spam-like manner. The infected machines can often be used as remote-controlled zombie computers, for more conventional spamming or DDoS attacks. Sometimes trojans are spammed to phish for bank account details, or to set up a pool of zombies without using a virus.

History

The term *spam* is derived from the Monty Python **SPAM sketch**, set in a cafe where everything on the menu includes SPAM luncheon meat. As the server recites the SPAM-filled menu, presently a chorus of Viking patrons drowns out all normal conversation with a song, repeating "SPAM, SPAM, SPAM, SPAM" and singing "lovely SPAM, wonderful SPAM" over and over again, stopping all conversation, hence SPAMming the dialogue. The excessive amount of SPAM in the sketch comes from British rationing in World War II. SPAM was one of the few foods that was not restricted and widely available, so by the time of the sketch, the British were fed up with the luncheon meat. Another similarity is that everything on the menu comes with SPAM, therefore representing that you can't order something without receiving something you don't want, much like one can't be active on the Internet and never have spam sent to your e-mail address(es).

Although the first known instance of unsolicited commercial e-mail occurred in 1978 (unsolicited electronic messaging had already taken place over other media, with the first recorded instance being via telegram on September 13, 1904), the term "spam" for this practice had not yet been applied. In the 1980s the term was adopted to describe certain abusive users who frequented BBSs and MUDs, who would repeat "SPAM" a huge number of times to scroll other users' text off the screen. In the early Chat rooms in services like PeopleLink and the early days of AOL, they actually flooded the screen with sizeable quotes from the Monty Python routine. This was generally used as a tactic by insiders of a particular group who wanted to drive newcomers out of the room so the usual conversation could continue. This act, previously termed *flooding* or *trashing*, came to be called *spamming* as well. By analogy, the term was soon applied to any large amount of text broadcast by one user, or sometimes by many users.

It later came to be used on Usenet to mean excessive multiple posting—the repeated posting of the same message. The first evident usage of this sense was by Joel Furr in the aftermath of the ARMM incident of March 31, 1993, in which a piece of experimental software released dozens of recursive messages onto the *news.admin.policy* newsgroup. Soon, this use had also become established—to spam Usenet was to flood newsgroups with junk messages.

Commercial spamming started in force on March 5, 1994, when a pair of lawyers, Laurence Canter and Martha Siegel, began using bulk Usenet posting to advertise immigration law services. The incident was commonly termed the "Green Card spam", after the subject line of the postings. The two went on to widely promote spamming of both Usenet and e-mail as a new means of advertisement—over the objections of Internet users they labeled "anti-commerce radicals." Within a few years, the focus of spamming (and antispam efforts) moved chiefly to e-mail, where it remains today.

There are three popular fake etymologies of the word "spam". The first, promulgated by Canter & Siegel themselves, is that "spamming" is what happens when one dumps a can of SPAM luncheon meat into a fan blade. The second is the backronym "**s**hit **p**osing **a**s **m**ail." The third is similar, using "**s**tupid **p**ointless **a**nnoying **m**essages."

Hormel Foods Corporation, the makers of SPAM® luncheon meat, do not object to the Internet use of the term "spamming." However, they do ask that the capitalized word "SPAM" be reserved to refer to their product and trademark. By and large, this request is obeyed in forums which discuss spam—to the extent that to write "SPAM" for "spam" brands the writer as a newbie. However, Hormel has begun to press the trademark issue—first, when a firm registered the trademark "SpamArrest" in 2003, Hormel sued to invalidate the mark, and more recently two failed attempts to revoke the mark "spambuster".

Alternate meanings

The term "spamming" is also used in the older sense of something repetitious and disruptive by players of various video games, most often first-person shooters or fighting games. For shooters, it refers to "area denial" tactics—repeatedly firing rockets or other explosive shells into an area—or to any tactic whereby a large volume of ammunition is expended in the hope of either scoring chance hits, covering teammates' advance with suppressive fire, or clearing or defending an area from an enemy presence. In fighting games, spamming most often refers to overuse of particularly powerful moves, especially if they are easy to execute.

Whether such tactics are viewed as cheating or abusive varies from game to game, community to community. Analogous to camping, the tactical advantage gained by those thus engaged is the crux of the issue. If every player defensively "spams", and no one makes the offensive push, there will be no opportunities for players to come into conflict, and thus there will be no game. Games like Capture the Flag help to break this deadlock by providing incentive to invade enemy territory, however risky.

Conversely, the same term may be used to describe those who flood the in-game chat with needlessly profuse and/or frequent messaging, similar to messaging spam mentioned

above. Although perceptions vary within the gaming community, in most arenas excessive messaging is unwelcome. On the other hand, in the role-playing games MUD, MUSH, and MUCK, players happily continue using the word in this original sense, with no implication of abuse. When a player returns to the terminal after a brief break to find his or her screen wonderfully filled with pages of random chat, it's still called "spam".

SPAM could also be taken to mean a set of humorous English backronyms, including: Short/Stupid/Silly Particularly/Pointless Annoying Messages, Self-Promotional Advertising Material, Self Propelled Automatic Mail, Send Post All Members, Sending Persistently Annoying Mail, and Shit Posing As Mail.

Costs of spam

The California legislature found that spam cost United States organizations alone more than $10 billion in 2004, including lost productivity and the additional equipment, software, and manpower needed to combat the problem.

Spam's direct effects include the consumption of computer and network resources, and the cost in human time and attention of dismissing unwanted messages. In addition, spam has costs stemming from the *kinds* of spam messages sent, from the *ways* spammers send them, and from the *arms race* between spammers and those who try to stop or control spam. In addition, there are the opportunity cost of those who forgo the use of spam-afflicted systems. There are the direct costs, as well as the indirect costs borne by the victims - both those related to the spamming itself, and to other crimes that usually accompany it, such as financial theft, identity theft, data and intellectual property theft, virus and other malware infection, child pornography, fraud, and deceptive marketing.

The methods of spammers are likewise costly. Because spamming contravenes the vast majority of ISPs' acceptable-use policies, most spammers have for many years gone to some trouble to conceal the origins of their spam. E-mail, Usenet, and instant-message spam are often sent through insecure proxy servers belonging to unwilling third parties. Spammers frequently use false names, addresses, phone numbers, and other contact information to set up "disposable" accounts at various Internet service providers. In some cases, they have used falsified or stolen credit card numbers to pay for these accounts. This allows them to quickly move from one account to the next as each one is discovered and shut down by the host ISPs.

The costs of spam also include the collateral costs of the struggle between spammers and the administrators and users of the media threatened by spamming.

Many users are bothered by spam because it impinges upon the amount of time they spend reading their e-mail. Many also find the content of spam frequently offensive, in that pornography is one of the most frequently advertised products. Spammers send their spam largely indiscriminately, so pornographic ads may show up in a work place e-mail inbox—or a child's, the latter of which is illegal in many jurisdictions. Recently, there has been a noticeable increase in spam advertising websites that contain child pornography.

Some spammers argue that most of these costs could potentially be alleviated by having spammers reimburse ISPs and individuals for their material. There are two problems with this logic: first, the rate of reimbursement they could credibly budget is not nearly high

enough to pay the direct costs; and second, the human cost (lost mail, lost time, and lost opportunities) is basically unrecoverable.

E-mail spam exemplifies a tragedy of the commons: spammers use resources (both physical and human), without bearing the entire cost of those resources. In fact, spammers commonly do not bear the cost at all. This raises the costs for everyone. In some ways spam is even a potential threat to the entire e-mail system, as operated in the past.

Since e-mail is so cheap to send, a tiny number of spammers can saturate the Internet with junk mail. Although only a tiny percentage of their targets are motivated to purchase their products (or fall victim to their scams), the low cost sometimes provides a sufficient conversion rate to keep spamming alive. Furthermore, even though spam appears not to be economically viable as a way for a reputable company to do business, it suffices for professional spammers to convince a tiny proportion of gullible advertisers that it is viable for those spammers to stay in business. Finally, new spammers go into business every day, and the low costs allow a single spammer to do a lot of harm before finally realizing that the business is not profitable.

Some companies and groups "rank" spammers; spammers who make the news are sometimes referred to by these rankings (Spamhaus' "TOP 10 spam service ISPs", The 10 Worst ROKSO Spammers). The necessary secretiveness of the operations makes uncertainty about how they actually determine "how bad" a spammer is unavoidable. Also, spammers may target different networks to different extents, depending on how successful they are at attacking the target. Thus considerable resources are employed to actually measure the amount of spam generated by a single person or group. For example, victims that use common antispam hardware, software or services provide opportunities for such tracking. Nevertheless, such rankings should be taken with a grain of salt.

To better understand the cost of spam to an organization, MX Logic Email Defense has posted a cost of spam calculator on their website.

Continuously updated statistics from postini track the ebb and flow of e-mail abuse without ranking spammers.

Political issues

Spamming remains a hot discussion topic. In fact, many online users have even suggested (though they were presumably joking) that cruel forms of capital punishment would be appropriate for spammers. In 2004, the seized Porsche of an indicted spammer was advertised on the internet; this revealed the extent of the financial rewards available to those who are willing to commit duplicitous acts online. However, some of the possible means used to stop spamming may lead to other side effects, such as increased government control over the Net, loss of privacy, barriers to free expression, and even the commercialization of e-mail.

One of the chief values favored by many long-time Internet users and experts, as well as by many members of the public, is the free exchange of ideas. Many have valued the relative anarchy of the Internet, and bridle at the idea of restrictions placed upon it. A common refrain from spam-fighters is that spamming itself abridges the historical freedom of the

Internet, by attempting to force users to carry the *costs* of material which they would not choose.

An ongoing concern expressed by parties such as the Electronic Frontier Foundation and the ACLU has to do with so-called "stealth blocking", a term for ISPs employing aggressive spam blocking without their users' knowledge. These groups' concern is that ISPs or technicians seeking to reduce spam-related costs may select tools which (either through error or design) also block non-spam e-mail from sites seen as "spam-friendly". SPEWS is a common target of these criticisms. Few object to the existence of these tools; it is their use in filtering the mail of users who are not informed of their use which draws fire.

Some see spam-blocking tools as a threat to free expression—and laws against spamming as an untoward precedent for regulation or taxation of e-mail and the Internet at large. Even though it is possible in some jurisdictions to treat some spam as unlawful merely by applying existing laws against trespass and conversion, some laws specifically targeting spam have been proposed. In 2004, United States passed the Can Spam Act of 2003 which provided ISPs with tools to combat spam. This act allowed Yahoo! to successfully sue Eric Head, reportedly one of the biggest spammers in the world, who settled the lawsuit for several thousand U.S. dollars in June 2004. But the law is criticized by many for not being effective enough. Indeed, the law was supported by some spammers and organizations which support spamming, and opposed by many in the antispam community. Examples of effective anti-abuse laws that respect free speech rights include those in the U.S. against unsolicited faxes and phone calls, and those in Australia and a few U.S. states against spam.

Court cases

Attorney Laurence Canter was disbarred by the Supreme Court of Tennessee in 1997 for sending prodigious amounts of spam advertising his immigration law practice.

Robert Soloway lost a case in a federal court against the operator of a small Oklahoma-based Internet service provider who accused him of spamming. In another case against Soloway, U.S. Judge Ralph G. Thompson granted a motion by plaintiff Robert Braver for a default judgment and permanent injunction against him. The judgment includes a statutory damages award of $10,075,000 under Oklahoma law.

In the first successful case of its kind, Mr. Nigel Roberts from the Channel Islands won £270 against Media Logistics UK who sent junk e-mails to his personal account.

Stopping e-mail abuse

E-mail has become the subject of much abuse, in the form of both spamming and E-mail worm programs. Both of these flood the in-boxes of E-mail users with junk E-mails, wasting their time and money, and often carrying offensive, fraudulent, or damaging content. This article describes the efforts being made to **stop E-mail abuse** and ensure that E-mail continues to be usable in the face of these threats.

Protection against spam

End users can protect themselves from the brunt of spam's impact in numerous ways.

Spam filters

The continuing increase in spam has resulted in rapid growth in the use of *spam filter* programs: software designed to examine incoming email and separate spam emails from genuine email messages intended for the user.

Unwanted e-mail can be filtered at the desktop, the network email server/email gateway, the Internet Service Provider's email gateway, or all three locations. While network managers and ISPs can choose hardened email security appliances, services or software designed to interdict both spam and viruses, desktop users are frequently limited to a software-based solution.

A number of commercial spam filtering programs exist and are readily available, but many freeware and shareware spam filters are also available for easy downloading and installation. Spam filters are currently included as standard features in nearly every available email client, though the quality of these built-in filters can be low; for some users, this may necessitate the use of a higher quality filtering solution.

Preventing Address Harvesting

Preventing spammers from obtaining your email address doesn't really solve the spam problem, any more than avoiding all but lowest crime areas of a city solves crime. Many people cannot hide their email addresses and most people want to meet new people via email. They just don't want the flood of spam. It may, however, reduce the amount of spam that you receive.

One way that spammers obtain email addresses to target is to trawl the Web and Usenet for strings which look like addresses, using a spambot. Contact forms and address munging are good ways to prevent email addresses from appearing on these forums. If the spammers can't find the address, the address won't get spam.

There are other ways that spammers can get addresses such as "dictionary attacks" in which the spammer generates a number of likely-to-exist addresses out of names and common words. For instance, if there is someone with the address `adam@example.com`, where 'example.com' is a popular ISP or mail provider, it is likely that he frequently receives spam.

Address munging
Posting anonymously, or with an entirely faked name and address, is one way to avoid this "address harvesting", but users should ensure that the faked address is not valid. Users who want to receive legitimate email regarding their posts or Web sites can alter their addresses in some way that humans can figure out but spammers haven't (yet). For instance, `joe@example.net` might post as `joeNOS@PAM.example.net`, or display his email address as an image instead of text. This is called *address munging,* from the jargon word "mung" meaning to break.

Contact Forms
Contact forms allow users to send email by filling out forms in a web browser. The web server takes the form data and forwards it to an email address. The user (and therefore the spam harvester) never sees the email address. Contact forms have the drawback that they require a website that supports server side scripts. They are also inconvenient to the message sender as he is not able to use his preferred e-mail client. Finally if the software used to run the contact forms is buggy or badly designed they can become spam tools in their own right.

Disposable e-mail addresses

Many email users sometimes need to give an address to a site without complete assurance that the site will not spam, or leak the address to spammers. One way to mitigate the risk of spam from such sites is to provide a *disposable* email address -- a temporary address which forwards email to your real account, but which you can disable or abandon whenever you see fit.

A number of services provide disposable address forwarding. Addresses can be manually disabled, can expire after a given time interval, or can expire after a certain number of messages have been forwarded. Some of these services allow easier creation of disposable addresses via various techniques.

Defeating Web bugs and JavaScript

Many modern mail programs incorporate Web browser functionality, such as the display of HTML, URLs, and images. This can easily expose the user to pornographic or otherwise offensive images in spam. In addition, spam written in HTML can contain JavaScript programs to direct the user's Web browser to an advertised page, or to make the spam message difficult or impossible to close or delete. In some cases, spam messages have contained attacks upon security vulnerabilities in the HTML renderer, using these holes to install spyware. (Some computer viruses are borne by the same mechanisms.) Also, the HTML can be used to signal whether a spam message is actually read and seen by a user.

Users can defend against these methods by using mail clients which do not automatically display HTML, images or attachments, or by configuring their clients not to display these by default.

Avoiding responding to spam

It is well established that some spammers regard responses to their messages -- even responses which say "Don't spam me" -- as confirmation that an email address refers validly to a reader. Likewise, many spam messages contain Web links or addresses which the user is directed to follow to be removed from the spammer's mailing list.

In several cases, spam-fighters have tested these links and addresses and confirmed that they do not lead to the recipient address's removal -- if anything, they lead to more spam.

In late 2003, the USA FTC launched a public relations campaign to encourage email users to simply *never respond to a spam email -- ever*. This campaign stemmed from the tendency of casual email users to reply to spam, in order to complain and request the spammer to cease sending spam.

Perhaps more significantly, since the sender address fields borne by spam messages are almost always forged, a reply to a spam message is likely to reach an innocent third party if it reaches anyone at all.

In Usenet, it is widely considered even more important to avoid responding to spam. Many ISPs have software that seeks out and destroys duplicate messages. Someone may see a spam and respond to it before it is cancelled by their server, which can have the effect of reposting the spammer's spam for them; since it is not just a duplicate, this reposted copy will last longer.

Reporting spam

The majority of ISPs explicitly forbid their users from spamming, and eject from their service users who are found to have spammed. Tracking down a spammer's ISP and reporting the offense often leads to the spammer's service being terminated. Unfortunately, it can be difficult to track down the spammer -- and while there are some online tools to assist, they are not always accurate. Also occasionally spammers own their own netblocks. In this case the abuse contact for the netblock can be the spammer itself and can confirm your address as live.

Examples of these online tools are SpamCop, Network Abuse Clearinghouse and Blue Frog. These provide automated or semi-automated means to report spam to ISPs. Some spam-fighters regard them as inaccurate compared to what an expert in the email system can do; however, most email users are not experts.

Consumers may also forward "unwanted or deceptive spam" to an email address (`spam@uce.gov`) maintained by the FTC. The database so collected is used to prosecute perpetrators of various types of scam or deceptive advertising.

Defense against email worms

In the past several years, scores of worm programs have used email systems as a conduit for infection. The worm program transmits itself in an email message, usually as a MIME attachment. In order to infect a computer, the executable worm attachment must be opened. In almost all cases, this means the user must click on the attachment. The worm also requires a software environment compatible with its programming.

Email users can defend against worms in a number of ways, including:

- Avoiding email client software which supports executable attachments. The most frequently-targeted client software for email worms is Microsoft Outlook and Outlook Express, both of which can easily be made to open executable attachments. However, other Windows-based email software is not immune to worms.
- Using an operating system which does not provide an environment compatible with present worms. Essentially all current email worms affect only the Microsoft Windows operating system. They cannot execute on Macintosh, Unix, GNU/Linux, or other operating systems. In some cases, it is conceivable that a worm could be written for one of these systems; however, various security features militate against it.

- Using up-to-date anti-virus software to detect incoming worms and quarantine or delete them before they can take effect.
- Being skeptical of unsolicited email attachments. Since worms and other email-borne malware arrive in this form, some email users simply refuse to open attachments that the sender has not given them advance notice of.

Examination of anti-spam methods

There are a number of services and software systems that mail sites and users can use to reduce the load of spam on their systems and mailboxes. Some of these depend upon rejecting email from Internet sites known or likely to send spam. Others rely on automatically analyzing the content of email messages and weeding out those which resemble spam. These two approaches are sometimes termed *blocking* and *filtering.*

Blocking and filtering each have their advocates and advantages. While both reduce the amount of spam delivered to users' mailboxes, blocking does much more to alleviate the bandwidth cost of spam, since spam can be rejected before the message is transmitted to the recipient's mail server. Filtering tends to be more thorough, since it can examine all the details of a message. Many modern spam filtering systems take advantage of machine learning techniques, which vastly improve their accuracy over manual methods. However, some people find filtering intrusive to privacy, and many mail administrators prefer blocking to deny access to their systems from sites tolerant of spammers.

DNSBLs

DNS-based Blackhole Lists, or DNSBLs, are used for heuristic filtering and blocking. A site publishes lists (typically of IP addresses) via the DNS, in such a way that mail servers can easily be set to reject mail from those sources. There are literally scores of DNSBLs, each of which reflects different policies: some list sites known to emit spam; others list open mail relays or proxies; others list ISPs known to support spam. Other DNS-based anti-spam systems list known good ("white") or bad ("black") IPs domains or URLs, including RHSBLs and URIBLs. For history, details, and examples of DNSBLs, see DNSBL.

Content-based filtering

Until recently, content filtering techniques relied on mail administrators specifying lists of words or regular expressions disallowed in mail messages. Thus, if a site receives spam advertising "herbal Viagra", the administrator might place these words in the filter configuration. The mail server would thence reject any message containing the phrase.

Content based filtering can also filter based on content other than the words and phrases that make up the body of the message. Primarily, this means looking at the header of the email, the part of the message that contains information about the message, and not the body text of the message. Spammers will often spoof fields in the header in order to hide their identities, or to try to make the email look more legitimate than it is; many of these spoofing methods can be detected. Also, spam sending software often produces a header that violates the RFC 2822 standard on how the email header is supposed to be formed.

Disadvantages of this static filtering are threefold: First, it is time-consuming to maintain. Second, it is prone to false positives. Third, these false positives are not equally distributed:

manual content filtering is prone to reject legitimate messages on topics related to products advertised in spam. A system administrator who attempts to reject spam messages which advertise mortgage refinancing may easily inadvertently block legitimate mail on the same subject.

Finally, spammers can change the phrases and spellings they use, or employ methods to try to trip up phrase detectors. This means more work for the administrator. However, it also has some advantages for the spam fighter. If the spammer starts spelling "Viagra" as "V1agra" or "Via_gra", it makes it harder for the spammer's intended audience to read their messages. If they try to trip up the phrase detector, by, for example, inserting an invisible-to-the-user HTML comment in the middle of a word ("Via<!---->gra"), this sleight of hand is itself easily detectable, and is a good indication that the message is spam. And if they send spam that consists entirely of images, so that anti-spam software can't analyze the words and phrases in the message, the fact that there *is* no readable text in the body can be detected.

However, content filtering can also be implemented by examining the URLs present (i.e. spamvertised) in an email message. This form of content filtering is much harder to disguise as the URLs must resolve to a valid domain name. Extracting a list of such links and comparing them to published sources of spamvertised domains is a simple and reliable way to eliminate a large percentage of spam via content analysis.

Statistical filtering

Statistical filtering was first proposed in 1998 by Mehran Sahami et al., at the AAAI-98 Workshop on Learning for Text Categorization. A statistical filter is a kind of document classification system, and a number of machine learning researchers have turned their attention to the problem. Statistical filtering was popularized by Paul Graham's influential 2002 article *A Plan for Spam*, which proposed the use of naive Bayes classifiers to predict whether messages are spam or not – based on collections of spam and nonspam ("ham") email submitted by users.

Statistical filtering, once set up, requires no maintenance per se: instead, users mark messages as spam or nonspam and the filtering software learns from these judgements. Thus, a statistical filter does not reflect the software author's or administrator's biases as to content, but it *does* reflect the *user's* biases as to content; a biochemist who is researching Viagra won't have messages containing the word "Viagra" flagged as spam, because "Viagra" will show up often in his or her legitimate messages. A statistical filter can also respond quickly to changes in spam content, without administrative intervention.

Spammers have attempted to fight statistical filtering by inserting many random but valid "noise" words or sentences into their messages while attempting to hide them from view, making it more likely that the filter will classify the message as neutral) Attempts to hide the noise words include setting them in tiny font or the same colour as the background. However, these noise countermeasures seem to have been largely ineffective.

Software programs that implement statistical filtering include Bogofilter, the e-mail programs Mozilla and Mozilla Thunderbird, and later revisions of SpamAssassin. Another

interesting project is CRM114 which hashes phrases and does bayesian classification on the phrases.

There is also the free mail filter POPFile which sorts mail in as many categories as you want (family, friends, co-worker, spam, whatever) with bayesian filtering.

Checksum-based filtering

Checksum-based filter takes advantage of the fact that often, for any individual spammer, all of the messages he or she sends out will be mostly identical, the only differences being web bugs, and when the text of the message contains the recipient's name or email address. Checksum-based filters strip out everything that might vary between messages, reduce what remains to a checksum, and look that checksum up in a database which collects the checksums of messages that email recipients consider to be spam (some people have a button on their email client which they can click to nominate a message as being spam); if the checksum is in the database, the message is likely to be spam.

The advantage of this type of filtering is that it lets ordinary users help identify spam, and not just administrators, thus vastly increasing the pool of spam fighters. The disadvantage is that spammers can insert unique invisible gibberish -- known as *hashbusters* -- into the middle of each of their messages, thus making each message unique and having a different checksum. This leads to an arms race between the developers of the checksum software and the developers of the spam-generating software.

Checksum based filtering methods include:

- Distributed Checksum Clearinghouse
- Vipul's Razor

Authentication and Reputation (A&R)

A number of systems have been proposed to allow acceptance of email from servers which have authenticated in some fashion as senders of only legitimate email. Many of these systems use the DNS, as do DNSBLs; but rather than being used to list nonconformant sites, the DNS is used to list sites authorized to send email, and (sometimes) to determine the reputation of those sites. Other methods of identifying ham and spam are still used. The A&R allows much ham to be more reliably identified, which allows spam detectors to be made more sensitive without causing more false positive results. The increased sensitivity allows more spam to be identified as such. Also, A&R methods tend to be less resource-intensive than other filtering methods, which can be skipped for messages identified by A&R as ham.

Sender-supported whitelists and tags

There are a small number of organizations which offer IP whitelisting and/or licensed tags that can be placed in email (for a fee) to assure recipients' systems that the messages thus tagged are not spam. This system relies on legal enforcement of the tag. The intent is for email administrators to whitelist messages bearing the licensed tag.

A potential difficulty with such systems is that the licensing organization makes its money by licensing more senders to use the tag -- not by strictly enforcing the rules upon

licensees. A concern exists that senders whose messages are more likely to be considered spam who would accrue a greater benefit by using such a tag. The concern is that these factors form a perverse incentive for licensing organizations to be lenient with licensees who have offended. However, the value of a license would drop if it was not strictly enforced, and financial gains due to enforcement of a license itself can providee an additional incentive for strict enforcement. The Habeas mail classing system attempts to further address this issue this by classing email according to origin, purpose, and permission. The purpose is to describe why the email is not likely spam, but permission based email.

Ham passwords

Another approach for countering spam is to use a "ham password". Systems that use ham passwords ask unrecognised senders to include in their email a password that demonstrates that the email message is a "ham" (not spam) message. Typically the email address and ham password would be described on a web page, and the ham password would be included in the "subject" line of an email address. Ham passwords are often combined with filtering systems, to counter the risk that a filtering system will accidentally identify a ham message as a spam message.

The "plus addressing" technique appends a password to the "username" part of the email address.

Cost-based systems

Since spam occurs primarily because it is so cheap to send, a proposed set of solutions require that senders pay some cost in order to send spam, making it uneconomic.

Stamps
Some gatekeeper such as Microsoft would sell electronic stamps, and keep the proceeds. Or a Micropayment, such as Electronic money would be paid by the sender to the recipient or their ISP, or some other gatekeeper.

Hashcash
Hashcash and simil ar systems require that a sender pay a computational cost by performing a calculation that the receiver can later verify. Verification must be much faster than performing the calculation, so that the computation slows down a sender but does not significantly impact a receiver. The point is to slow down machines that send most of spam -- often millions and millions of them. While every user that wants to send email to a moderate number of recipients suffers just a seconds' delay, sending millions of emails would take an unaffordable amount of time.

Bonds
As a refinement to stamp systems was the idea of requiring that the micropayment only be retained if the recipient considered the email to be abusive. This addressed the principal objection to stamp systems: popular free legitimate mailing list hosts would be unable to continue to provide their services if they had to pay postage for every message they sent out.

Issues
A difficulty that must be dealt with by most anti-spam methods, including DNSBLs, Authentication and Reputation (A&R), Sender-supported whitelists and tags, Ham passwords, cost-based systems, Heuristic filtering, and Challenge/response systems is that spammers already (illegally) use other people's computers to send spam. The computers in question are already infected with viruses and spyware operated by the spam senders, in some cases seriously damaging the computer's responsiveness to the legitimate user. Spam from the legitimate user's computer can be sent using the user's and/or system's identity, list of correspondents, reputation, credentials, stamps, hashcash and/or bonds. The added motivation to steal from such systems in order to abuse these things may simply impel spammers to infect more computers and cause greater damage. On the other hand, this could compel computer users to finally secure their systems, reducing Botnets, which would have myriad other benefits, as they are used for extortion, phishing, and terorrism, as well as spam. Ultimately, any system that holds senders responsible for the mail they send needs to deal with the situation of irresponsible senders that may send both spam and ham.

Heuristic filtering

Heuristic filtering, such as is implemented in the program SpamAssassin, uses some or all of the various tests for spam mentioned above, and assigns a numerical score to each test. Each message is scanned for these patterns, and the applicable scores tallied up. If the total is above a fixed value, the message is rejected or flagged as spam. By ensuring that no single spam test by itself can flag a message as spam, the false positive rate can be greatly reduced.

Tarpits and Honeypots

A *tarpit* is any server software which intentionally responds pathologically slowly to client commands. A *honeypot* is a server which attempts to attract attacks. Some mail administrators operate tarpits to impede spammers' attempts at sending messages, and honeypots to detect the activity of spammers. By running a tarpit which appears to be an open mail relay, or which treats acceptable mail normally and known spam slowly, a site can slow down the rate at which spammers can inject messages into the mail facility.

One tarpit design is the *teergrube*, whose name is simply German for "tarpit." This is an ordinary SMTP server which intentionally responds very slowly to commands. Such a system will bog down SMTP client software, as further commands cannot be sent until the server acknowledges the earlier ones. Several SMTP MTAs, including Postfix and Exim, have a teergrube capacity built-in: when confronted with a client session which causes errors such as spam rejections, they will slow down their responding. A similar approach is taken by TarProxy.

Another design for tarpits directly controls the TCP/IP protocol stack, holding the spammer's network socket open without allowing any traffic over it. By reducing the TCP window size to zero, but continuing to acknowledge packets, the spammer's process may be tied up indefinitely. This design is more difficult to implement than the former. Aside from anti-spam purposes, it has also been used to absorb attacks from network worms.

As of late 2005 much of the spam sent is through so-called "zombie" systems, of which there are potentially a very large number. This makes the actual effectiveness of tarpits questionable, as there are so many spam sources that slowing just a few has little real effect on the volume of spam received.

Another approach is simply an imitation MTA (open relay honeypot) which gives the appearance of being an open mail relay. Spammers who probe systems for open relay will find such a host and attempt to send mail through it, wasting their time and potentially revealing information about themselves and the source of spam to the unexpected alert entity (in comparison to the anticipated careless or unskilled operator typically in charge of open relay MTA systems) that operates the honeypot. Such a system may simply discard the spam attempts, submit them to DNSBLs, or store them for analysis. It may be possible to examine or analyze the intercepted spam to find information that allows other countermeasures. (One honeypot operator was able to alert a freemail supplier to a large number of accounts that had been created as dropboxes for the receipt of responses to spam. Disabling these dropbox email accounts made the entire spam run, including the spam messages relayed through actual open relays, useless to the spammer: he could not receive any of the responses to the spam sent by gullible customers.) The SMTP honeypot may also selectively deliver relay test messages to give a stronger appearance of open relay (though care is needed here as this means the honeypot itself and the network it is on could end up on spam blacklists). SMTP honeypots of this sort have been suggested as a way that end-users can interfere with spammers' activities (code: Java, Python).

As of late 2005 open relay abuse to send spam has greatly declined, resulting in a lowered active effectiveness of open relay honeypots. (Passively, the honeypots or threat of same create an inducement for spammers to not abuse open relays.) Other types of honeypot (below) may still have great effectiveness.

Spammers also abuse open proxies, and open proxy honeypots (proxypots) have had substantial success. Ron Guillmette reported in 2003 that he succeeded in getting over 100 spammer accounts terminated in under 3 months, using his network (of unspecified size) of proxypots. At that time spammers were so careless that they sent spam directly from their servers to the abused open proxy, making determination of the identity of the spammer's IP address trivial so that it was easy to report the spammer to the ISP in control of that IP address and easy for that ISP to terminate the spammer's account.

Unlike most other anti-spam techniques tarpits and honeypots work at the relay, proxy, or zombie (collectively, "abuse") level. They work by targeting spammer behavior rather than targeting spam content. One beneficial fallout from this is that these tools are not required to have any means of distinguishing spam from non-spam. Because they capture spam at the abuse level they are not part of any legitimate email pathway and it can be confidently assumed that what they capture is 100% spam or spam-related (e.g., test messages.) Anti-spam measures at (or after) the destination server level protect specific email addresses but must include code to distinguish spam from non-spam. Anti-spam measures at the abuse level protect whatever the email addresses are that are being targeted by the spam directed through them and are hence non-specific but need no code to distinguish spam from non-spam. The main purpose of abuse-level tools is targeting spam and spammers themselves while the main purpose of server-level tools is to protect speecific email

addresses. What abuse-level tools lose in specificity may be more than made up by the inherent simplicity that results from not having to be able to separate valid email from invalid email.

In late 2005 Microsoft announced that it had converted an actual zombie system to a zombie honeypot. One result of this was a lawsuit by Microsoft against about 20 defendants, based on evidence collected by the zombie honeypot.

Note that there is some terminological confusion. Some people refer to "spamtraps" as "honeypots." In this context a "spamtrap" is an email address created specifically to attract spam. These run at the destination level rather than at the relay, proxy or "spam zombie" level.

Challenge/response systems

Another method which may be used by internet service providers (or by specialized services) to combat spam is to require unknown senders to pass various tests before their messages are delivered. These strategies are termed **challenge/response systems** or **C/R**, are currently controversial among email programmers and system administrators.

e-Mail spam

E-mail spam is a subset of spam that involves sending nearly identical messages to thousands (or millions) of recipients. Perpetrators of such spam ("spammers") often harvest addresses of prospective recipients from Usenet postings or from web pages, obtain them from databases, or simply guess them by using common names and domains. By popular definition, spam occurs without the permission of the recipients.

Overview

As the recipient directly bears the cost of delivery, storage, and processing, one could regard spam as the electronic equivalent of "postage-due" junk mail. However, the Direct Marketing Association will point to the existence of "legitimate" e-mail marketing. Most commentators classify e-mail-based marketing campaigns where the recipient has "opted in" to receive the marketer's message as "legitimate".

Spammers frequently engage in deliberate fraud to send out their messages. Spammers often use false names, addresses, phone numbers, and other contact information to set up "disposable" accounts at various Internet service providers. They also often use falsified or stolen credit card numbers to pay for these accounts. This allows them to move quickly from one account to the next as the host ISPs discover and shut down each one.

Spammers frequently go to great lengths to conceal the origin of their messages. They do this by spoofing e-mail addresses (much easier than Internet protocol spoofing). The e-mail protocol (SMTP) has no authentication by default, so the spammer can easily make a message appear to originate from any e-mail address. To prevent this, some ISPs and domains require the use of SMTP-AUTH, allowing positive identification of the specific account from which an e-mail originates.

Spammers cannot completely spoof e-mail delivery chains (the 'Received' header), since the receiving mailserver records the actual connection from the last mailserver's IP address. To counter this, some spammers forge additional delivery headers to make it appear as if the e-mail had previously traversed many legitimate servers. But even when the fake headers are identified, tracing an e-mail message's route is usually fruitless. Many ISPs have thousands of customers, and identifying spammers is tedious and generally not considered worth the effort.

Spammers frequently seek out and make use of vulnerable third-party systems such as open mail relays and open proxy servers. The SMTP system, used to send e-mail across the Internet, forwards mail from one server to another; mail servers that ISPs run commonly require some form of authentication that the user is a customer of that ISP. Open relays, however, do not properly check who is using the mail server and pass all mail to the destination address, making it quite a bit harder to track down spammers.

Increasingly, spammers use networks of virus-infected Windows PCs (zombies) to send their spam. Zombie networks are also known as Botnets.

Spoofing can have serious consequences for legitimate e-mail users. Not only can their e-mail inboxes get clogged up with "undeliverable" e-mails in addition to volumes of spam, they can mistakenly be identified as a spammer. Not only may they receive irate e-mail from spam victims, but (if spam victims report the e-mail address owner to the ISP, for example) their ISP may terminate their service for spamming.

Legality

Sending spam violates the Acceptable Use Policy (AUP) of almost all Internet Service Providers, and can lead to the termination of the sender's account. Many jurisdictions, such as the United States of America, which regulates via the CAN-SPAM Act of 2003, regard spamming as a crime or as an actionable tort.

Article 13 of the European Union Directive on Privacy and Electronic Communications (2002/58/EC) provides that the EU member states shall take appropriate measures to ensure that unsolicited communications for the purposes of direct marketing are not allowed either without the consent of the subscribers concerned or in respect of subscribers who do not wish to receive these communications, the choice between these options to be determined by national legislation.

Accessing privately owned computer resources without the owner's permission counts as illegal under computer crime statutes in most nations. Deliberate spreading of computer viruses is also illegal in the United States and elsewhere.

Thus, some of spammers' most common behaviors are criminal quite independently of the legal status of spamming per se. Even before the advent of laws specifically banning or regulating spamming, spammers have been successfully prosecuted under computer fraud and abuse laws for wrongfully using others' computers.

Avoiding spam

Computer users can avoid e-mail spam in several ways:

- End-users can use automated e-mail filtering on their own computers.
- System administrators can use appropriate tools to trap e-mail spam at the mail server level, either by use of software or special appliances.
- Spam can be reported to appropriate ISP so that the spamming can be stopped.
- By giving out one's ISP e-mail address only to closely trusted acquaintances, friends, and relatives, and using web based e-mail services for everyone else.
- By ensuring that those acquaintances, friends and relatives who have been trusted with one's e-mail address do not include the person who wants to avoid spam's e-mail address in the "To" or "CC" fields when sending several copies of an e-mail to ensure that, when such e-mails are forwarded, to avoid one's e-mail addresses from appearing in an ammassing list of e-mail addresses
- By creating a unique e-mail address for each person or site you wish to communicate with. This can be done using an online mail forwarding service, or with administrative access to your own e-mail server. If spam is received on one of these addresses, you immediately know who leaked or sold your address to spammers, and you can also cancel the affected e-mail address.
- End-users can take precautions to avoid needlessly publicising their e-mail addresses or protect them from e-mail harvesting by spam bots, such as by using e-mail forms that do not display the address in the webpage code, or by address munging.
- Using anti-virus and anti-spyware programs with regularly updated definitions to avoid having their computers hijacked and used as spammer tools.
- Users are also advised to configure their e-mail clients to disable rich content features such as HTML mail and automatic downloading of images. Downloaded images can be used by spammers to identify valid e-mail addresses.
- By periodically performing an internet search for one's own email address, and if necessary getting the appropriate website administrator to remove it.

Anti-spam programmers have released several tools—intended for both end users and for systems administrators—which automate the highlighting, removal or filtering of e-mail spam by scanning through incoming and outgoing e-mails in search of traits typical of spam. Modern anti-spam systems are usually very effective at protecting you from spam.

Like other forms of theft, spam should be reported to the appropriate people so that it can be stopped. Services, such as spamcop, make this easy to do. While this may not immediately decrease the amount of spam you receive significantly, it will reduce the amount of spam that everyone receives.

The best way to avoid spam involves avoiding making one's e-mail address available to spammers, directly or indirectly.

Basic computer literacy should include an understanding of the basics of spamming and spam avoidance. One should never reply to a spam e-mail, or click an "opt-out" link (this

simply confirms that an e-mail address is "live"). Users should not reveal their e-mail addresses on porn, warez and other shady sites.

If a web site requests registration in order to allow useful operations, such as posting in Internet forums, a user may give a temporary disposable address—set up and used only for such a purpose—periodically deleting such temporary e-mail accounts from their e-mail servers. (Users should notify such forums of the new replacement addresses if they wish to continue interaction for valid purposes.) For example, free services such as spamgourmet.com and spamhole.com allow a user to create a temporary e-mail address which forwards e-mail to you for a set period of time, and then becomes invalid.

Avoiding sending spam

Anti-spam ISPs and technicians have published a number of resources to help systems and e-mail users avoid sending spam inadvertently or through misunderstanding the e-mail system — such as the MAPS Guidelines for Mailing List Management. These guidelines aim to help legitimate users of bulk e-mail who wish not only to comply with anti-spam laws, but also to avoid appearing to customers or Internet partners as spammers.

Broadly, such guidelines promote the idea that e-mail recipients must grant permission before others may send them bulk e-mail. In effect, senders must not send bulk e-mail to users who have not *opted in* to receive it. This contrasts with the view of e-mail promoted by many bulk e-mailers, who claim that senders should feel free to send to any user who has not *opted out* of receiving it.

Many spammers, however, do not even comply with an opt-out régime. Although U.S. and other laws require that commercial e-mailers cease sending to recipients who have opted out, many spam messages contain fraudulent opt-out instructions. In some cases, spammers have used the opt-out function as a way of confirming that someone actually read a spam message. In 2004 some spam messages turned out to contain malware for Microsoft Windows which victims triggered by clicking an opt-out link.

Cost-based methods

A number of persons have proposed "e-mail postage" systems, under which e-mail senders would be required to pay money, perform a resource-intensive computation, or post a bond, for each message sent. Proponents include Microsoft's Bill Gates. The intention of e-mail postage is to deter spam by making it too expensive to send a large number of messages.

However, since spammers already use other people's computers to spam, there is every reason to believe that they would offload the postage charge onto others as well.

Confirmed opt-in

One difficulty occurs in implementing opt-in mailing lists: many means of gathering user e-mail addresses remain susceptible to forgery. For instance, if a company puts up a Web form to allow users to subscribe to a mailing list about its products, a malicious person can enter other people's e-mail addresses — to harass them, or to make the company appear to be spamming. (To most anti-spammers, if the company sends e-mail to these forgery victims, it *is* spamming, albeit inadvertently.)

To prevent this abuse, MAPS and other anti-spam organizations encourage that all mailing lists use **confirmed opt-in** (also known as *verified opt-in* and (by spammers themselves) as *double opt-in*). That is, whenever an address is presented for subscription to the list, the list software should send a confirmation message to that address. The confirmation message contains no advertising content, so it is not construed to be spam itself — and the address is not added to the list unless the recipient responds to the confirmation message.

All modern mailing list management programs (such as GNU Mailman, Majordomo, and qmail's ezmlm) support confirmed opt-in by default.

Highest Amount of Spam Received
The owner of the domain name acme.com is currently thought to hold the record, receiving over one million spam emails per day. This is in comparision with Microsoft founder Bill Gates who, according to Steve Ballmer, receives four million per year.

How spammers operate

Gathering of addresses

In order to send spam, spammers need to obtain the e-mail addresses of the intended recipients. Toward this end, both spammers themselves and *list merchants* gather huge lists of potential e-mail addresses. Since spam is, by definition, unsolicited, this *address harvesting* is done without the consent (and sometimes against the expressed will) of the address owners. As a consequence, spammers' address lists are remarkably inaccurate. A single spam run may target tens of millions of possible addresses -- many of which are invalid, malformed, or undeliverable.

Spam differs from other forms of direct marketing in many ways, one of them being that it costs no more to send to a larger number of recipients than a smaller number. For this reason, there is little pressure upon spammers to limit the number of addresses targeted in a spam run, or to restrict it to persons likely to be interested. One consequence of this fact is that many people receive spam written in languages they cannot read — a good deal of spam sent to English-speaking recipients is in Chinese or Korean, for instance. Likewise, lists of addresses sold for use in spam frequently contain malformed addresses, duplicate addresses, and addresses of role accounts such as `postmaster`.

Spammers may harvest e-mail addresses from a number of sources. A popular method uses e-mail addresses which their owners have published for other purposes. Usenet posts, especially those in archives such as Google Groups, frequently yield addresses. Simply searching the Web for pages with addresses — such as corporate staff directories — can yield thousands of addresses, most of them deliverable. Spammers have also subscribed to discussion mailing lists for the purpose of gathering the addresses of posters. The DNS and WHOIS systems require the publication of technical contact information for all Internet domains; spammers have illegally trawled these resources for e-mail addresses. Many spammers utilize programs called web spiders to find e-mail addresses on web pages.

A recent, controversial tactic, called *"e-pending"*, involves the *appending* of *e-mail* addresses to direct-marketing databases. Direct marketers normally obtain lists of prospects from sources such as magazine subscriptions and customer lists. By searching the Web and other

resources for e-mail addresses corresponding to the names and street addresses in their records, direct marketers can send targeted spam e-mail. However, as with most spammer "targeting", this is imprecise; users have reported, for instance, receiving solicitations to mortgage their house at a specific street address — with the address being clearly a business address including mail stop and office number.

Spammers sometimes use various means to confirm addresses as deliverable. For instance, including a Web bug in a spam message written in HTML may cause the recipient's mail client to transmit the recipient's address, or any other unique key, to the spammer's Web site.

Likewise, spammers sometimes operate Web pages which purport to remove submitted addresses from spam lists. In several cases, these have been found to subscribe the entered addresses to receive more spam.

Delivering spam messages

Internet users and system administrators have deployed a vast array of techniques to block, filter, or otherwise banish spam from users' mailboxes. Almost all Internet service providers forbid the use of their services to send spam or to operate spam-support services. Both commercial firms and volunteers run subscriber services dedicated to blocking or filtering spam, such as AppRiver, Brightmail, Postini, and the various DNSBLs.

Using Webmail services

A common practice of spammers is to create accounts on free webmail services, such as Hotmail, to send spam or to receive e-mailed responses from potential customers. Because of the amount of mail sent by spammers, they require several e-mail accounts, and use web bots to automate the creation of these accounts.

In an effort to cut down on this abuse, many of these services have adopted a system called the captcha: users attempting to create a new account are presented with a graphic of a word, which uses a strange font, on a difficult to read background. Humans are able to read these graphics, and are required to enter the word to complete the application for a new account, while computers are unable to get accurate readings of the words using standard OCR techniques. Blind users of captchas typically get an audio sample.

Spammers have, however, found a means of circumventing this measure. Reportedly, they have set up sites offering free pornography: to get access to the site, a user displays a graphic from one of these webmail sites, and must enter the word. Once the bot has successfully created the account, the user gains access to the pornographic material.

Using other people's computers

Early on, spammers discovered that if they sent large quantities of spam directly from their ISP accounts, recipients would complain and ISPs would shut their accounts down. Thus, one of the basic techniques of sending spam has become **to send it from someone else's computer** and network connection. By doing this, spammers protect themselves in several ways: they hide their tracks, get others' systems to do most of the work of delivering messages, and direct the efforts of investigators towards the other systems rather than the spammers themselves. The increasing broadband usage gave rise to a great number of

computers that are online as long as they are turned on, and whose owners do not always take steps to protect them from malware. A botnet consisting of several hundred compromised machines can effortlessly churn out millions of messages per day. This also complicates the tracing of spammers.

Open relays

In the 1990s, the most common way spammers did this was to use **open mail relays**. An open relay is an MTA, or mail server, which is configured to pass along messages sent to it from any location, to any recipient. In the original SMTP mail architecture, this was the default behavior: a user could send mail to practically any mail server, which would pass it along towards the intended recipient's mail server.

The standard was written in an era before spamming when there were few hosts on the internet, and those on the internet abided by a certain level of conduct. While this cooperative, open approach was useful in ensuring that mail was delivered, it was vulnerable to abuse by spammers -- and abused it soon was. Spammers could forward batches of spam through open relays, leaving the job of delivering the messages up to the relays.

In response, mail system administrators concerned about spam began to demand that other mail operators configure MTAs to cease being open relays. The first DNSBLs, such as MAPS RBL and the now-defunct ORBS, aimed chiefly at allowing mail sites to refuse mail from known open relays.

Open proxies

Within a few years, open relays became rare and spammers resorted to other tactics, most prominently the use of **open proxies**. A *proxy* is a network service for making indirect connections to other network services. The client connects to the proxy and instructs it to connect to a server. The server perceives an incoming connection from the proxy, not the original client. Proxies have many purposes, including Web-page caching, protection of privacy, filtering of Web content, and selectively bypassing firewalls.

An *open* proxy is one which will create connections for *any* client to *any* server, without authentication. Like open relays, open proxies were once relatively common, as many administrators did not see a need to restrict access to them.

A spammer can direct an open proxy to connect to a mail server, and send spam through it. The mail server logs a connection from the proxy -- not the spammer's own computer. This provides an even greater degree of concealment for the spammer than an open relay, since most relays log the client address in the headers of messages they pass. Open proxies have also been used to conceal the sources of attacks against other services besides mail, such as Web sites or IRC servers.

Besides relays and proxies, spammers have used other insecure services to send spam. One example is the now-infamous `FormMail.pl`, a CGI script to allow Web-site users to send e-mail feedback from an HTML form. Several versions of this program, and others like it, allowed the user to redirect e-mail to arbitrary addresses. Spam sent through **open**

FormMail scripts is frequently marked by the program's characteristic opening line: "Below is the result of your feedback form."

As spam from proxies and other "spammable" resources grew, DNSBL operators started listing their IP addresses, as well as open relays.

Spammer viruses

In 2003, spam investigators saw a radical change in the way spammers sent spam. Rather than searching the global network for exploitable services such as open relays and proxies, spammers began creating "services" of their own. By commissioning computer viruses designed to deploy proxies and other spam-sending tools, spammers could harness hundreds of thousands of end-user computers.

Most of the major Windows e-mail viruses of 2003, including the Sobig and Mimail virus families, functioned as **spammer viruses**: viruses designed expressly to make infected computers available as spamming tools.

Besides sending spam, spammer viruses serve spammers in other ways. Beginning in July 2003, spammers started using some of these same viruses to perpetrate distributed denial-of-service (DDoS) attacks upon DNSBLs and other anti-spam resources. Although this was by no means the first time that illegal attacks have been used against anti-spam sites, it was perhaps the first wave of *effective* attacks.

In August of that year, engineering company Osirusoft ceased providing DNSBL mirrors of the SPEWS and other blocklists, after several days of unceasing attack from virus-infected hosts. The very next month, DNSBL operator Monkeys.com succumbed to the attacks as well. Other DNSBL operators, such as Spamhaus, have deployed global mirroring and other anti-DDoS methods to resist these attacks.

Obfuscating message content

Many spam-filtering techniques work by searching for patterns in the headers or bodies of messages. For instance, a user may decide that all e-mail she receives with the word "Viagra" in the subject line is spam, and instruct her mail program to automatically delete all such messages. To defeat such filters, the spammer may intentionally misspell commonly-filtered words, use Leetspeak or insert other characters, as in the following examples:

- V1agra
- Via'gra
- V I A G R A
- Vaigra
- \ /iagra
- Vi@graa

The principle of this method is to leave the word readable to humans (whose pattern-recognition skills make them remarkably adept at picking out the true meaning of

misspelled words), but not recognizable to a literally-minded computer program. This is effective up to a point. Eventually, filter patterns become generic enough to recognize the word "Viagra" no matter how misspelled -- or else they target the obfuscation methods themselves, such as insertion of punctuation into unusual places in a word.

(Note: Using most common variations, it is possible to spell "Viagra" in at least 1,300,925,111,156,286,160,896 ways.)

HTML-based e-mail gives the spammer more tools to obfuscate text. Inserting HTML comments between letters can foil some filters, as can including text made invisible by setting the font color to white on a white background, or shrinking the font size to the smallest fine print.

Another common ploy involves presenting the text as an image, which is either sent along or loaded from a remote server. This can be foiled by not permitting an e-mail-program to load images.

As Bayesian filtering has become popular as a spam-filtering technique, spammers have started using methods to weaken it. To a rough approximation, Bayesian filters rely on word probabilities. If a message contains many words which are only used in spam, and few which are never used in spam, it is likely to be spam. To weaken Bayesian filters, some spammers now include lines of irrelevant, random words alongside the sales pitch. A variant on this tactic may be borrowed from the Usenet abuser known as "Hipcrime" -- to include passages from books taken from Project Gutenberg, or nonsense sentences generated with "dissociated press" algorithms. Randomly generated phrases can create spamoetry (spam poetry) or spam art.

Another method used to masquerade spam as legitimate messages is the use of autogenerated sender names in the **From:** field, ranging from realistic ones such as "Jackie F. Bird" to (either by mistake or intentionally) bizarre attention-grabbing names such as "Sloppiest U. Epiglottis" or "Attentively E. Behavioral". Return addresses are also routinely auto-generated.

Spam-support services

A number of other online activities and business practices are considered by anti-spam activists to be connected to spamming. These are sometimes termed **spam-support services**: business services, other than the actual sending of spam itself, which permit the spammer to continue operating. Spam-support services can include processing orders for goods advertised in spam, hosting Web sites or DNS records referenced in spam messages, or a number of specific services as follows:

Some Internet hosting firms advertise **bulk-friendly** or **bulletproof hosting**. This means that, unlike most ISPs, they will not terminate a customer for spamming. These hosting firms operate as clients of larger ISPs, and many have eventually been taken offline by these larger ISPs as a result of complaints regarding spam activity. Thus, while a firm may advertise bulletproof hosting, it is ultimately unable to deliver without the connivance of its upstream ISP. However, some spammers have managed to get what is called a *pink contract* (see below) — a contract with the ISP that allows them to spam without being disconnected.

A few companies produce **spamware**, or software designed for spammers. Spamware varies widely, but may include the ability to import thousands of addresses, to generate random addresses, to insert fraudulent headers into messages, to use dozens or hundreds of mail servers simultaneously, and to make use of open relays. The sale of spamware is illegal in eight U.S. states.

So-called **millions CDs** are commonly advertised in spam. These are CD-ROMs purportedly containing lists of e-mail addresses, for use in sending spam to these addresses. Such lists are also sold directly online, frequently with the false claim that the owners of the listed addresses have requested (or "opted in") to be included. Such lists often contain invalid addresses.

A number of DNSBLs, including the MAPS RBL, Spamhaus SBL, and SPEWS, target the providers of spam-support services as well as spammers.

Related vocabulary

Unsolicited commercial e-mail (UCE)

- The most common type of spam, e-mails sent to recipients who did not request them, promoting a commercial service that makes money for the spammer.

Unsolicited bulk e-mail (UBE)

- E-mail viruses (worms) sent by infected computers. Also forwarded hoaxes (e.g. virus warnings), political advocacy spam, and chain letters sent by a person to many other people.

Pink contract

- A service contract offered by an ISP which offers bulk e-mail service to spamming clients, in violation of that ISP's publically posted acceptable use policy. Not used by reputable ISPs (if they want to remain reputable).

Spamvertised

- Adjective that describes a website "advertised" by spammers.

Spam bait

Spam bait is e-mail sent in the hopes that the unwitting recipient will reply, indicating to the original sender that the recipient's e-mail address is a valid one and can be added to a mailing list for spam.

Spam bait can be hard to counter. The most effective strategy is a negative one: do not reply to such messages!

Another, more technical strategy is to ensure that your e-mail account does not automatically fetch embedded images in webmail. Some e-mail contains HTML code whose main purpose is to provide feedback to the spam bait sender. When your e-mail software requests delivery of the image, the address of that image can contain a tracking code. The

sender returns the requested image but also gleans from the tracking code the fact that they've found another valid e-mail address. Then you get more spam.

Word salad

Word salad is a mixture of seemingly meaningful words that together signify nothing; the phrase draws its name from the positive symptom of psychosis, Word salad (mental health). When applied to a physical theory, "word salad" it is a derogatory description that labels the theory as senseless or utterly devoid of meaning.

In the context of computer science and linguistics, explicitly constructed word salad is a tool for demonstrating the difference between random utterance and coherent expression of thought. Software such as the Dissociated Press within emacs demonstrates the construction of interesting-but-meaningless word salad from large samples of coherent language, by constructing new, random documents that share some of the same word or letter clustering properties as the language sample. These word salads appear as natural language to the inattentive eye or ear, but are clearly meaningless when read or listened to with full attention. In the 21st century, e-mail spammers have begun using word salad construction as a way to elude e-mail filtering.

In spam e-mail

In response to the growing problem of spam e-mail, filtering tools became available starting around 2002 which implemented a widely employed method known as the naive Bayes classifier. This method uses the probability of various words appearing in spam emails to automatically classify them as spam. For a short time, this worked fairly well to classify emails as probable spam. In response, spammers developed **word salad** to fool programs employing this method of classification. By adding large amounts of random text somewhere in their message, spammers hope to confuse Bayesian classifiers into classifying the message as "ham e-mail" (non-spam e-mail). Typically, this text contains random words from a dictionary.

Algorithms for detecting word salad are clearly possible and not particularly difficult to implement. They would be, for the most part, more computationally intensive than most rules used by spam filters today (2006). A statistical approach based on Zipf's law of word frequency has potential in detecting simple word salad, as do grammar checking and the use of natural language processing algorithms. Statistical Markovian analysis, where short phrases are used to determine if they are likely to occur in normal English sentences, is another statistical approach that would be effective against completely random phrasing but might be fooled by Dissociated Press techniques.

Sentence and paragraph salad

In a related technique, actual text from some large corpus of legitimate English (the plays of Shakespeare, other etexts distributed by Project Gutenberg, random world wide web pages, or the like) is added into the email. This approach attempts to get around algorithms that could be devised to detect the more primitive form of word salad.

Paragraph salad will reduce the effectiveness of any of the algorithms mentioned above and will lead to higher scores with any Bayesian filters. The only algorithms that might thwart sentence and paragraph salad would be very high level and expensive natural language processing, some kind of artificial intelligence algorithm involving a search engine, or exhaustive listing of spam emails. All of these techniques would be exceptionally expensive, and would likely not be very successful at filtering spam despite their high cost.

Letter salad

On an even smaller scale than word salad, spammers use misspellings of words to try to thwart Bayesian filters. Misspelling Viagra as Via6ra, \/|/\Gr/\, or any one of a number of other ways, or even using characters from international character sets is an attempt to avoid the high efficiency with which a Bayesian filter would classify any email containing certain words as spam. A simple spell checker might significantly reduce the effectiveness of letter salad approaches, yet most present spam filters do not use one.

The lengths to which some spammers have gone with letter salad have often produced illegible, almost laughable messages. Reading such email has become akin to deciphering complex custom license plates.

Future

As spam filters get better at detecting simple word and letter salad, spammers will likely migrate towards sentence and paragraph salad techniques. In the process of obscuring their message from improving spam filters, they will also obscure their message from potential targets of their advertising, virus distribution, or phishing. At some point, the profitability of spam may be brought down to the point that its volume is substantially reduced.

Recommendations

End users should take no action upon receiving email with word salad content, or whose sender or purpose is unclear. Opening questionable email, and especially clicking on links contained in it, may risk overall information security.

Spamvertising

Spamvertising is the practice of sending E-mail spam, advertising a website. In this case, it is a portmanteau of the words "spam" and "advertising".

It also refers to vandalizing wikis, blogs and online forums with hyperlinks in order to get a higher search engine ranking for the vandal's website. Spamvertisers insert links to their websites (typically, sites purporting to sell some commercial product) and add keywords of common or related searches. The apparent goal is that a search engine will find the vandalized page full of links and improve the popularity rating of the pages to which they link. This is typically done by automated editing programs which look for editable text fields in web forms and automatically fill them in with web links. The links typically lead to pills, porn and poker sites.

This practice has led to many editable online resources employing anti-spam countermeasures including the use of captchas to attempt to prevent automated editing.

DNSBL

A **DNS-based Blackhole List** (**DNSBL**, **Real-time Blackhole List** or **RBL**), is a means by which an Internet site may publish a list of IP addresses, in a format which can be easily queried by computer programs on the Internet. As the name suggests, the technology is built on top of the Internet DNS or Domain Name System. DNSBLs are chiefly used to publish lists of addresses linked to spamming. Most mail transport agent (mail server) software can be configured to reject or flag messages which have been sent from a site listed on one or more such lists.

DNSBL names a medium, not any specific list or policy. There has been a good deal of controversy over the past several years over the operation of specific lists, such as the MAPS RBL, ORBS, and SPEWS.

History of DNSBLs

The first DNSBL was the Real-time Blackhole List (RBL), created in 1997 by Paul Vixie as part of his Mail Abuse Prevention System (MAPS). Vixie, an influential Internet programmer and administrator, encouraged the authors of sendmail and other mail software to implement RBL clients. These allowed the mail software to query the RBL and reject mail from listed sites. However, the purpose of the RBL was not simply to block spam—it was to educate Internet service providers and other Internet sites about spam and related problems, such as open SMTP relays. Before an address would be listed on the RBL, volunteers and MAPS staff would attempt repeatedly to contact the persons responsible for it and get its problems corrected.

Soon after the advent of the RBL, others started developing their own lists with different policies. One of the first was Alan Brown's Open Relay Behavior-modification System (ORBS). This used automated testing to discover and list mail servers running as open mail relays—exploitable by spammers to carry their spam. ORBS was controversial at the time because many people felt running an open relay was acceptable, and that scanning the Internet for open mail servers could be abusive.

In recent events (2003), a number of DNSBLs have come under denial-of-service attacks. Since no party has admitted to these attacks nor been discovered responsible, their purpose is a matter of speculation. However, many observers believe the attacks are perpetrated by spammers in order to interfere with the DNSBLs' operation or hound them into shutting down. In August 2003, the firm *Osirusoft*, an operator of several DNSBLs including one based on the SPEWS data set, shut down its lists after suffering weeks of near-continuous attack.

A number of parties, such as the Electronic Frontier Foundation and Peacefire, have raised concerns about some use of DNSBLs by ISPs. One joint statement issued by a group

including EFF and Peacefire addressed "stealth blocking", in which ISPs use DNSBLs or other spam-blocking techniques without informing their clients.

DNSBL Operation

To operate a DNSBL requires three things: a domain to host it under, a nameserver for that domain, and a list of addresses to publish.

It is possible to serve a DNSBL using BIND, the popular DNS software. However, BIND is inefficient for zones containing large numbers of addresses, particularly DNSBLs which list entire Classless Inter-Domain Routing netblocks. DNSBL-specific software—such as Michael J. Tokarev's rbldnsd or Daniel J. Bernstein's rbldns—is faster, uses less memory, and is easier to configure than the general-purpose BIND. Alternatively, Simplicita Software offers a commercial DNSBL server that provides additional benefits such as point-in-time auditing and 24/7 IP address monitoring.

The hard part of operating a DNSBL is populating it with addresses. DNSBLs intended for public use usually have specific, published policies as to what a listing means, and must be operated accordingly to attain or keep public confidence.

DNSBL Queries

When a mail server receives a connection from a client, and wishes to check that client against a DNSBL (let's say, *spammers.example.net*), it does more or less the following:

- Take the client's IP address—say, *192.168.42.23*—and reverse the bytes, yielding *23.42.168.192*.
- Append the DNSBL's domain name: *23.42.168.192.spammers.example.net*.
- Look up this name in the DNS as a domain name ("A" record). This will return either an address, indicating that the client is listed; or an "NXDOMAIN" ("No such domain") code, indicating that the client is not.
- Optionally, if the client is listed, look up the name as a text record ("TXT" record). Most DNSBLs publish information about why a client is listed as TXT records.

Looking up an address in a DNSBL is thus similar to looking it up in reverse-DNS. The differences are that a DNSBL lookup uses the "A" rather than "PTR" record type, and uses a forward domain (such as *spammers.example.net* above) rather than the special reverse domain *in-addr.arpa*.

There is an informal protocol for the addresses returned by DNSBL queries which match. Most DNSBLs return an address in the 127.0.0.0/8 IP loopback network. The address 127.0.0.2 indicates a generic listing. Other addresses in this block may indicate something specific about the listing—that it indicates an open relay, proxy, spammer-owned host, etc.

DNSBL Policies

Different DNSBLs have different policies. DNSBL policies differ from one another on three fronts:

- **Goals.** What does the DNSBL *seek* to list? Is it a list of open-relay mail servers or open proxies—or of IP addresses known to send spam—or perhaps of IP addresses belonging to ISPs that harbor spammers?
- **Nomination.** How does the DNSBL *discover* addresses to list? Does it use nominations submitted by users? Spam-trap addresses or honeypots?
- **Listing lifetime.** How long does a listing *last*? Are they automatically expired, or only removed manually? What can the operator of a listed host do to have it delisted?

Terminology

The proprietary term *RBL* is sometimes erroneously used in place of the generic *DNSBL*. RBL is a service mark of MAPS LLC. Some pieces of mail software have configuration parameters for the use of "RBLs" or "RBL domains", used to set the DNSBLs that the software should use. This may be trademark dilution.

Note: Trend Micro bought MAPS LLC in June 2005.

An *RHSBL* or *Right-Hand-Side Blackhole List* is a DNSBL which lists domain names rather than IP addresses. The term comes from the "right-hand side" of an email address -- the part after the @ sign -- which clients look up in the RHSBL.

Criticisms

Email users who find their messages blocked from mail servers that use DNSBLs often object vociferously, sometimes to the extent of attacking the existence of the lists themselves. The following lists are controversial:

- Lists of dynamic and dial-up IP addresses. Some mail sites choose not to accept messages from dynamic addresses, since they are often home computers exploited by spammer viruses. This can inconvenience users who wish to run their own mail servers on residential ISP connections or local MTAs on laptops for example.
- Lists that include "spam-support operations", such as MAPS RBL. A spam-support operation is a site that may not directly send spam, but provides commercial services for spammers, such as hosting of Web sites that are advertised in spam. Refusal to accept mail from spam-support operations is intended as a boycott to encourage such sites to cease doing business with spammers, at the expense of inconveniencing non-spammers who use the same site as spammers.
- Predictive ("early warning") lists, notably SPEWS. SPEWS lists addresses belonging to spam-support operations, under the hypothesis that such addresses are more likely to send spam in the future. SPEWS "escalates" listings, increasing the size of the netblock listed, as a site continues to support spam.

Although many have voiced objections to specific DNSBLs, few people object to the principle that mail-receiving sites should be able to reject undesired mail systematically. One who does is John Gilmore, who deliberately operates an open mail relay. Gilmore accuses DNSBL operators of violating antitrust law.

For Joe Blow to refuse emails is legal (though it's bad policy, akin to "shooting the messenger"). But if Joe and ten million friends all gang up to make a blacklist, they are exercising illegal monopoly power.

Spammers have pursued lawsuits against DNSBL operators on similar grounds. In 2003, a newly-formed corporation calling itself "EmarketersAmerica" filed suit against a number of DNSBL operators in Florida court. Backed by spammer Eddy Marin, the company claimed to be a trade organization of "email marketers" and that DNSBL operators Spamhaus and SPEWS were engaged in restraint of trade. The suit was eventually dismissed for lack of standing.

The Abusive Hosts Blocking List

The **Abusive Hosts Blocking List** is an abuse tracking and filtering system developed by The Summit Open Source Development Group, and based on the original Summit Blocking List (2000-2002).

DNSbl and RHSbl Lists

The AHBL operates several DNSbl lists and one RHSbl list for use in various types of services. While most of the data is automatically added and removed, the AHBL prefers to manually manage certain categories by hand for accuracy.

The **DNSbl** list was developed for use in SMTP services and is "a real time blocking system. This means that data is collected from various sources 24 hours a day, 7 days a week in real time, and merged into our database."The data includes spam sources, open proxies, open relays, DDoS drones, Usenet spam sources, and the controversial Shoot On Sight listing policy.

The **IRCbl** list is a reduced version of the DNSbl that does not include spam sources or other data unnecessary for use in IRC networks and other chat systems.

The **RHSbl** list is domain based rather than ip4r. It includes domains owned and/or operated by spammers, known abusive domains, and domains that are not used to send e-mail (on request of the domain owner). It is commonly used to block domains in the From: address of e-mail, as well as SURBL type systems that scan the links in e-mail.

The **TORbl** list is an ip4r based list of Tor (anonymity network) nodes. It includes only public exit nodes on the tor network.

Controversy

Several of the AHBL's actions since its creation have led to harsh criticism from other members of the spam fighting community for being overly aggressive and unreasonable. One such example is the complete blocking of Spain's largest Internet service provider, Telefonica.es, for more then 6 months "because of the ever increasing amount of spam and illegal 419 coming from rima-tde.net IP space."The AHBL has also publicly spoken out against the Spamhaus .Mail ICANN proposal.

The second major source of controversy comes from the AHBL's Shoot On Sight listing policy, which is commonly used by its administrators to force ISPs to take action against known abusers. It also tries to hold accountable individuals, companies, and providers responsible for legal threats and actions against spam fighters.

AHBL In Court

Richard Scoville/FreeSpeechStore vs. AHBL/SOSDG/Bruns/Kirch

After constant lawsuit threats for years prior to 2005, on December 17th, 2005 Richard Scoville of FreeSpeechStore.com (pro-se) sued the AHBL, SOSDG, Brian Bruns, and Andrew D. Kirch in Bexar County, TX for $3.525 million USD, claiming damage for personal humiliation, embarrassment, emotional distress, and other damages relating to his business (known as FreeSpeechStore).

Word of the lawsuit spread through various Usenet groups including NANAE after Scoville tried to publicly grandstand on his side of the case. Shortly after, Andrew D. Kirch released a public statement confirming the lawsuit, and Brian Bruns made a public plea for donations to help cover the cost of the defense.

In what the AHBL described as an attempt to intimidate potential donors, Scoville threatened to sue any individual or company that donated to the AHBL Legal Defense Fund. This caused the exact opposite of the desired effect, and several news sites carried the story of the lawsuit. Within one week, more then $4000 USD had been donated to the fund.

Due to jurisdictional concerns, the AHBL was granted a special appearance hearing, represented by Mary Claire Fischer (Attorney At Law, Bexar County, TX).

On January 5th 2006, Scoville provided witnesses to support his claim that the jurisdiction was valid in Bexar County, TX, which included the designers of his website, Brandon Zumwalt of Internet Concepts (his former hosting provider), and Detective Brian Padier of the San Antonio Police Department Computer Crimes Unit.

On January 6th 2006, the case was dismissed with prejudice.

The AHBL's administrators continue to hold the opinion that this lawsuit is frivolous, brought against them for the purpose of intimidation (to force removal of Scoville's website IP address from the AHBL's DNSbl) and to cause both Kirch and Bruns financial harm (through the expendature of large amounts of funds to defend the out of state cases).

Mr. Scoville continues to post libelious statements in teaser form on usenet via Google Groups to entice people to his pay-to-read website FreeSpeechStore.com (FSS) where free speech costs $4.00 per speech to read. Many of his FSS rants is his labelling a person as a pedophile or sexual deviant. He targets people well outside his local legal jurisdiction (San Antonio, Texas) such as Ottawa, Canada and Australia. A simple Google search can verify these facts.

Open mail relay

An **open mail relay** is an SMTP (e-mail) server configured in such a way that it allows anyone on the Internet to relay (i.e. send) e-mail through it.

History and technology

Until the 1990s this was the normal configuration for a mail server and was often the default on UNIX systems at installation. This was due, in part, to the traditional method in which e-mail (through and beyond the Internet) was passed from computer to computer via modems on telephone lines, often never touching the small Internet of the time. It was cheaper and simpler for e-mail to be passed from computer to computer until it reached its destination than to connect directly to the target computer (e.g. via modem) and log in to transfer the mail. Filtering and speed of e-mail delivery were not priorities. Furthermore, the small number of servers on the Internet generated a peer pressure on standards and conventions of its use. Moreover, on the government and educational servers with which the Internet was started, the transfer of commercial messages was forbidden by federal edict, as it was a government operation.

Nowadays, e-mail transfer by "relaying," or pass-along methods, is almost forgotten. Backbone networks and Internet switches make it cost effective and expeditious for end-user PCs or even cellphones to send mail directly to the target host, without need for relaying through a "middleman" site. The underlying communication methods of the Internet already provide end-to-end connectivity via a pass-along method.

Abuse by spammers

In the mid-1990s, once the Internet became more of a publicly available (and moreover, commercial) service, it wasn't long until it was utilized by mass-marketers, today known in the electronic world as spammers. As spam soon became widely unpopular, especially by e-mail server administrators who had to deal with the increased unsolicited traffic, spammers discovered that they could minimize or avoid detection by re-routing their e-mail through third party e-mail servers. After this practice became widespread, the mere practice of operating an open relay e-mail server became frowned upon among a significant number of Internet server administrators and other prominent users, many of whom were veterans of the Internet's non-commercial era.

Anti-spam efforts against open relays

Many ISPs use DNSBLs (DNS -based Blocking Lists) to disallow mail from open relays. Once a mail server is detected or reported that allows third parties to send mail through them, they will be added to one or more such lists, and other e-mail servers using those lists will reject any mail coming from those sites.

This trend reduced the percentage of mail senders that were open relays from over 90% down to well under 1% over several years. This led to spammers adopting other techniques, such as the use of open proxies to send spam. Although open relays are no longer widely used to send spam, many sites continue to refuse mail traffic from them.

One consequence of the new unacceptability of open relays was an inconvenience for some end users and certain internet service providers. To allow customers to use their e-mail addresses at Internet locations other than the company's systems (such as at school or work), many mail sites explicitly allowed open relaying so that customers could read and send e-mail via the ISP from any location. Once open relay became unacceptable due to abuse (and unusable due to blocking of open relays) ISPs and other sites had to adopt new protocols to allow remote users to send mail. These include SMTP-AUTH, POP-before-SMTP, and the use of virtual private networks (VPNs).

The Can Spam Act of 2003 makes it illegal to send spam through an open relay, but makes no provision regarding sending personal e-mail through them or regarding their operation.

Modern-day proponents

The most famous open mail relay operating today is probably that of John Gilmore, who argues that running an open relay is a free speech issue. His server is included on many open relay blacklists (many of which are generated by "automatic detection", that is, by anti-spam blacklisters sending an (unsolicited) test e-mail to other servers to see if they will be relayed). He has never sent any spam personally, yet these measures cause much of his outgoing e-mail to be blocked. Along with his further deliberate configuration of the server, his open relay enables people to send e-mail without their IP address being directly visible to the recipient and thereby send e-mail anonymously.

Gilmore contends he has a right to configure his computer however he pleases, and others have the right to configure their computers to ignore him. However, since open-relay blacklisting is most commonly done at the ISP level, many end users have this decision made for them without their explicit request. Many ISPs have been unwilling to remove the blacklists that prevent his e-mails from reaching recipients on the ISP's network or implement any other method (such as a whitelist) to allow his e-mail through. As a result, he is unable to communicate by e-mail with many of his friends and business partners.

Messaging spam

Messaging spam, sometimes called **SPIM**, is a type of spam where the target is instant messaging services.

The increase in messaging spam may be motivated by its rise in popularity as well as the many steps to crack down on spamming since the late 1990s.

Instant-messaging applications

Instant messaging (IM) systems, such as Yahoo! Messenger, AIM, MSN Messenger and ICQ, are popular targets for spammers. Many IM systems offer a directory of users, including demographic information such as age and sex. Advertisers can gather this information, sign on to the system, and send unsolicited messages.

Using privacy options to guard against messaging spam

To combat SPIM, many users choose to receive IMs only from people already on their contact list.

- With Yahoo Messenger, users can click Messenger -> Preferences -> Ignore List and check the box "Ignore anyone who is not on my Messenger List."
- With AOL's Instant Messenger, or AIM, users can click My AIM -> Edit Options -> Edit Preferences -> Privacy and check "Allow only users on my buddy list"
- With MSN Messenger, users can click Tools -> Options -> Privacy and check the box "Only people on my Allow List can see my status and send me messages"
- With BitWise IM, users can click Preferences -> Server / Contact List -> and check the box "Whitelist my contact list so that only users on my contact list can see me online or contact me"
- With Trillian, users can click Trillian -> Trillian Preferences -> Identities & Connections, and in all ICQ accounts click on the Miscellaneous tab and uncheck "Allow users to see my status on the web". There is also a plug-in available called Trillian Spam Challenge which asks unknown contacts from any medium to enter a phrase proving they're not a spambot, before the message is allowed through.

If you send someone you know an instant message, and they do not respond, it may be because they do not yet have you in their messenger's allow list. Ask the user to add you to their allow list first.

Using AIM 'warn' feature

The free AOL Instant Messenger (AIM) service allows users to 'warn' other users. The warning decreases the number of messages an account can send, slowing down spam, and shows the AIM address as warned to other users that it may try to message. This feature does have the potential for abuse, although such cases are minimal.

Windows messaging spam

In 2002, a number of spammers began using the Microsoft Windows Messaging service to get their message across. This isn't the same as the IM system "Windows Messenger"; rather, it is a function of Windows designed to allow servers to send alerts to administrator workstations. Windows Messaging spam appears as normal dialog boxes containing the spammer's message. Windows Messaging spam can be delivered using any NetBIOS port, so to block it at a firewall entails closing down ports 135 through 139, and 445.

Alternatively, Windows users can simply disable the messenger service entirely though the Windows services list available via Run/services.msc.

Messenger service spam, in particular, has lent itself to spammer use in a particularly circular scheme. In many cases, messenger spammers send messages to vulnerable Windows machines consisting of text like: "Annoyed by these messages? Visit this site." The link leads to a Web site where, for a fee, users are told how to disable the Windows messenger service. Though the messenger service is easily disabled for free by the user,

this scam works because it creates a perceived need and then offers an immediate solution. Often, the only "annoying messages" the user is receiving through messenger are advertisements to disable messenger itself.

Mobile phone spam

Mobile phone spam is a form of spamming directed at the text messaging service of a mobile phone. It is described as mobile spamming, sms spam, but is most frequently referred to as m-spam.

In 2002 and 2003, frequent users of cell phone text messages began to see an increase in the number of unsolicited (and generally unwanted) commercial advertisements being sent to their cell phones through text messaging.

In the United States, this use is regulated by the CAN-SPAM Act of 2003 and the Telephone Consumer Protection Act of 1991.

Often these messages consist of a simple request to call a number. Normal mobile phone etiquette often results in the call being returned by the user. When they then return the call, they are unaware that they have been fraudulently induced to call a premium-rate line. There is frequently an attempt to get them to hold on the line for as long as possible in order to maximise revenue from this fraud.

Another form of mobile phone fraud is the one-ring fraud, where an incoming call to a mobile phone is timed such that it will ring once (or without any sound at all), and then cut off before the user can answer. This leaves the missed call number on their phone, and the rest of the fraud is as above. In this case, it is the (real or apparent) calling number details which are being spammed to the phone, as these calls are made in their hundreds of thousands by autodialers at little or no cost to the originator, as there is no charge for calls which do not connect.

Both of these frauds can be combined with other frauds such as the advance fee fraud, as they act as a pre-screening stage for fraudsters to capture the telephone numbers of particularly trusting individuals.

As with spamming in general, there are usually no special laws against mobile phone spamming. However, existing laws can often be used to combat the problem. On June 10, 2004 Russian SMS spammer Dmitry Androsov was convicted for mobile phone spam by Chelyabinsk court. Dmitry was sued by Megafon, one of the largest Russian mobile operators, for sending SMS with expletives to more than 16000 mobile phone users. The fact that the sending was carried out using a Perl script allowed the court to convict him under the article 273 of the Russian Criminal Code (creation, use and distribution of malicious programs) to 1 year probational sentence and 3000 ruble (more than $100) fine.

On 12 April 2006, Singapore authorities has imposed a SGD$150,000 fine to mTouche content provider for unauthorised Use of End User Information and Unsolicited, Chargeable SMSes.

Newsgroup spam

Newsgroup spam is a type of spam where the targets are Usenet newsgroups.

Spamming of Usenet newsgroups actually pre-dates e-mail spam. The first widely recognized Usenet spam (though not the most famous) was posted on January 18, 1994 by Clarence L. Thomas IV, a sysadmin at Andrews University. Entitled "Global Alert for All: Jesus is Coming Soon", it was a fundamentalist religious tract claiming that "this world's history is coming to a climax." The newsgroup posting bot Serdar Argic also appeared in early 1994, posting tens of thousands of messages to various newsgroups, consisting of identical copies of a political screed relating to the Armenian Genocide.

The first *commercial* Usenet spam, and the one which is often (mistakenly) claimed to be the first Usenet spam of any sort, was an advertisement for legal services entitled "Green Card Lottery - Final One?". It was posted in April 1994 by Arizona lawyers Laurence Canter and Martha Siegel, and hawked legal representation for United States immigrants seeking papers ("green cards").

Usenet convention defines spamming as *excessive multiple posting,* that is, the repeated posting of a message (or substantially similar messages). During the early 1990s there was substantial controversy among Usenet system administrators (news admins) over the use of cancel messages to control spam. A *cancel message* is a directive to news servers to delete a posting, causing it to be inaccessible to those who might read it. Some regarded this as a bad precedent, leaning towards censorship, while others considered it a proper use of the available tools to control the growing spam problem.

A culture of neutrality towards content precluded defining spam on the basis of advertisement or commercial solicitations. The word "spam" was usually taken to mean **excessive multiple posting (EMP)**, and other neologisms were coined for other abuses — such as "velveeta" (from the processed cheese product) for *excessive cross-posting*. A subset of spam was deemed **cancellable spam**, for which it is considered justified to issue third-party cancel messages.

In the late 1990s, spam became used as a means of vandalising newsgroups, with malicious users committing acts of sporgery to make targeted newsgroups all but unreadable without heavily filtering. A prominent example occurred in alt.religion.scientology. Another known example is the Meow Wars.

The prevalence of Usenet spam led to the development of the Breidbart Index as an objective measure of a message's "spamminess". The use of the BI and spam-detection software has led to Usenet being policed by anti-spam volunteers, who purge newsgroups of spam by sending cancels and filtering it out on the way into servers. This very active form of policing has meant that Usenet is a far less attractive target to spammers than it used to be, and most of the industrial-scale spammers have now moved into e-mail spam instead.

Spit (VoIP spam)

SPIT (**SPam** over Internet Telephony) is a notional term devised by marketers to describe a problem that does not yet exist, but for which solutions may be sold. No examples have yet been seen in the wild, but Qovia of Frederick, Maryland, have recently filed two patent applications for technology to thwart spit. However, companies providing next generation telephone services to consumers are concerned about VoIP security. Kayote Networks Inc. provides a range of turnkey interconnect services overcoming SPIT.

Spyware

In the field of computing, the term **spyware** refers to a broad category of malicious software designed to intercept or take partial control of a computer's operation without the informed consent of that machine's owner or legitimate user. While the term taken literally suggests software that surreptitiously monitors the user, it has come to refer more broadly to software that subverts the computer's operation for the benefit of a third party.

Spyware differs from viruses and worms in that it does not usually self-replicate. Like many recent viruses, however, spyware – by design – exploits infected computers for commercial gain. Typical tactics furthering this goal include delivery of unsolicited pop-up advertisements; theft of personal information (including financial information such as credit card numbers); monitoring of Web- browsing activity for marketing purposes; or routing of HTTP requests to advertising sites.

As of 2005, spyware has become one of the pre-eminent security threats to computer-systems running Microsoft Windows operating-systems (and especially to users of Internet Explorer because of that browser's collaboration with the Windows operating system). Some malware on the Linux and Mac OS X platforms has behavior similar to Windows spyware, but to date has not become anywhere near as widespread.

History and development

The first recorded use of the term *spyware* occurred on October 17, 1994 in a Usenet post that poked fun at Microsoft's business model. Spyware later came to refer to espionage equipment such as tiny cameras. However, in early 2000 the founder of Zone Labs, Gregor Freund, used the term in a press release for the ZoneAlarm Personal Firewall. Since then, computer-users have used the term in its current sense.

In early 2000, Steve Gibson of Gibson Research realized that advertising software had been installed on his system, and he suspected that the software was stealing his personal information. After analyzing the software he determined that they were adware components from the companies Aureate (later Radiate) and Conducent. He eventually rescinded his claim that the ad software collected information without the user's knowledge, but still chastised the ad companies for covertly installing the spyware and making it difficult to remove.

As a result of his analysis in 2000, Gibson released the first anti-spyware program, OptOut, and many more software-based antidotes have appeared since then. International Charter now offers software developers a Spyware-Free Certification program.

According to a November 2004 study by AOL and the National Cyber-Security Alliance, 80% of surveyed users' computers had some form of spyware, with an average of 93 spyware components per computer. 89% of surveyed users with spyware reported that they did not know of its presence, and 95% reported that they had not given permission for the installation of the spyware.

Ads and malware

There is also class of advertising methods which may be considered unethical and perhaps even illegal. These include external applications which alter system settings (such as a browser's home page), spawn pop-ups, and insert advertisements into non-affiliated webpages. Such applications are usually labeled as spyware or adware. They may mask their questionable activities by performing a simple service, such as displaying the weather or providing a search bar. Some programs are effectively trojans. These applications are commonly designed so as to be difficult to remove or uninstall. The ever-increasing audience of online users, many of which are not computer-savvy, frequently lack the knowledge and technical ability to protect themselves from these programs.

Spyware, "adware", and tracking

The term *adware* frequently refers to any software which displays advertisements, whether or not it does so with the user's consent. Programs such as the Eudora mail client display advertisements as an alternative to shareware registration fees. These classify as "adware" in the sense of advertising-supported software, but not as spyware. They do not operate surreptitiously or mislead the user.

Many of the programs frequently classified as spyware function as *adware* in a different sense: their chief observed behavior consists of displaying advertising. Claria Corporation's Gator Software and Exact Advertising's BargainBuddy provide examples of this sort of program. Visited Web sites frequently install Gator on client machines in a surreptitious manner, and it directs revenue to the installing site and to Claria by displaying advertisements to the user. The user experiences a large number of pop-up advertisements.

Other spyware behaviors, such as reporting on websites the user visits, frequently accompany the displaying of advertisements. Monitoring web activity aims at building up a marketing profile on users in order to sell "targeted" advertisement impressions. The prevalence of spyware has cast suspicion upon other programs that track Web browsing, even for statistical or research purposes. Some observers describe the Alexa Toolbar, an Internet Explorer plug-in published by Amazon.com, as spyware (and some anti-spyware programs report it as such) although many users choose to install it.

Routes of infection

Spyware does not directly spread in the manner of a computer virus or worm: generally, an infected system does not attempt to transmit the infection to other computers. Instead, spyware gets on a system through deception of the user or through exploitation of software vulnerabilities.

The most direct route by which spyware can infect a computer involves the user installing it. However, users tend not to install software if they know that it will disrupt their working environment and compromise their privacy. So many spyware programs deceive the users, either by piggybacking on a piece of desirable software, or by tricking the users to do something that installs the software without them realizing. Recently, spyware has come to

include "rogue anti-spyware" programs, which masquerade as security software while actually doing damage.

Classically, a Trojan horse, by definition, smuggles in something dangerous in the guise of something desirable. Some spyware programs get spread in just this manner. The distributor of spyware presents the program as a useful utility — for instance as a "Web accelerator" or as a helpful software agent. Users download and install the software without immediately suspecting that it could cause harm. For example, Bonzi Buddy, a spyware program targeted at children, claims that:

He will explore the Internet with you as your very own friend and sidekick! He can talk, walk, joke, browse, search, e-mail, and download like no other friend you've ever had! He even has the ability to compare prices on the products you love and help you save money! Best of all, he's FREE!

Spyware can also come bundled with shareware or other downloadable software, as well as music CDs. The user downloads a program (for instance, a music program or a file-trading utility) and installs it, and the installer additionally installs the spyware. Although the desirable software itself may do no harm, the bundled spyware does. In some cases, spyware authors have paid shareware authors to bundle spyware with their software, as with the *Gator* spyware now marketed by Claria. In other cases, spyware authors have repackaged desirable free software with installers that add spyware.

A third way of distributing spyware involves tricking users by manipulating security features designed to prevent unwanted installations. The Internet Explorer Web browser, by design, prevents websites from initiating an unwanted download. Instead, a user action (such as clicking on a link) must normally trigger a download. However, links can prove deceptive: for instance, a pop-up ad may appear like a standard Windows dialog box. The box contains a message such as "Would you like to optimize your Internet access?" with links which look like buttons reading *Yes* and *No*. No matter which "button" the user presses, a download starts, placing the spyware on the user's system. Later versions of Internet Explorer offer fewer avenues for this attack.

Some spyware authors infect a system by attacking security holes in the Web browser or in other software. When the user navigates to a Web page controlled by the spyware author, the page contains code which attacks the browser and forces the download and install of spyware. The spyware author would also have some extensive knowledge of commercially-available anti-virus and firewall software. This has become known as a "drive-by download", which leaves the user a hapless bystander to the attack. Common browser exploits target security vulnerabilities in Internet Explorer and in the Microsoft Java runtime.

The installation of spyware frequently involves Microsoft's Internet Explorer. As the most popular Web browser, and with an unfortunate history of security issues, it has become the largest target. Its deep integration with the Windows environment and its scriptability make it an obvious point of attack into Microsoft Windows operating systems. Internet Explorer also serves as a point of attachment for spyware in the form of browser helper objects, which modify the browser's behavior to add toolbars or to redirect traffic.

In a few cases, a worm or virus has delivered a payload of spyware. For instance, some attackers used the W32.Spybot.Worm worm to install spyware that popped up pornographic ads on the infected system's screen. By directing traffic to ads set up to channel funds to the spyware authors, they can profit even by such clearly illegal behavior.

Effects and behaviors

Many Internet Explorer add-on toolbars monitor the user's activity. When installed and run without the user's consent, such add-ons count as spyware. Here multiple toolbars (including both spyware and innocuous ones) overwhelm an Internet Explorer session.

A piece of spyware rarely "lives" alone: an affected computer can rapidly become infected with large numbers of spyware components. Users frequently notice unwanted behavior and degradation of system performance. A spyware infestation can create significant unwanted CPU activity, disk usage, and network traffic—slowing down legitimate uses of these resources. Stability issues—application or system crashes—are also common. Spyware which interferes with the networking software commonly causes difficulty connecting to the Internet.

When Microsoft Windows users seek technical support—whether from computer manufacturers, Internet service providers, or other sources—spyware infection emerges as the most common cause. In many cases, the user has no awareness of spyware and assumes that the system performance, stability, and/or connectivity issues relate to hardware, to Microsoft Windows installation problems, or to a virus. Some owners of badly infected systems resort to buying an entire new computer system because the existing system "has become too slow". Badly infected systems may require a clean reinstall of all their software in order to restore the system to working order. This can become a time-consuming task, even for experienced users.

Only rarely does a single piece of software render a computer unusable. Rather, a computer rarely has only one infection. As the 2004 AOL study noted, if a computer has any spyware at all, it typically has dozens of different pieces installed. The cumulative effect, and the interactions between spyware components, typically cause the stereotypical symptoms reported by users: a computer which slows to a crawl, overwhelmed by the many parasitic processes running on it. Moreover, some types of spyware disable software firewalls and anti-virus software, and/or reduce browser security settings, thus opening the system to further opportunistic infections, much like an immune deficiency disease. Documented cases have also occurred where a spyware program disabled other spyware programs installed by its competitors.

Some other types of spyware (Targetsoft, for example) modify system files to make themselves harder to remove. (Targetsoft modifies the "Winsock" Windows Sockets files. The deletion of the spyware-infected file "inetadpt.dll" will interrupt normal networking usage.) Unlike users of many other operating systems, a typical Windows user has administrator-level privileges on the system, mostly for the sake of convenience. Because of this, any program which the user runs (intentionally or not) has unrestricted access to the system. Spyware, along with other threats, has led some Windows users to move to other

platforms such as Linux or Apple Macintosh, which such malware targets far less frequently.

Advertisements

Many spyware programs reveal themselves visibly by displaying advertisements. Some programs simply display pop-up ads on a regular basis—for instance, one every several minutes, or one when the user opens a new browser window. Others display ads in response to specific sites that the user visits. Spyware operators present this feature as desirable to advertisers, who may buy ad placement in pop-ups displayed when the user visits a particular site. It is also one of the purposes for which spyware programs gather information on user behavior.

Pop-up advertisements lead to some of users' most common complaints about spyware. A computer can become overwhelmed downloading or displaying ads. An infected computer rarely has only one spyware component installed—they more often number in the dozens—and so while a single program might display ads only infrequently, the cumulative effect becomes overwhelming.

Many users complain about irritating or offensive advertisements as well. As with many banner ads, many spyware advertisements use animation or flickering banners designed to catch the eye—thus they become highly visually distracting. Pop-up ads for pornography often display indiscriminately, including when children use the computer—possibly in violation of anti-pornography laws.

A further issue in the case of some spyware programs has to do with the replacement of banner ads on viewed web sites. Spyware that acts as a web proxy or a Browser Helper Object can replace references to a site's own advertisements (which fund the site) with advertisements that instead fund the spyware operator. This cuts into the margins of advertising-funded Web sites.

"Stealware" and affiliate fraud

A few spyware vendors, notably WhenU and 180 Solutions, have written what the New York Times has dubbed "stealware", and what spyware-researcher Ben Edelman terms *affiliate fraud*, also known as click fraud. These redirect the payment of affiliate marketing revenues from the legitimate affiliate to the spyware vendor.

Affiliate marketing networks work by tracking users who follow an advertisement from an "affiliate" and subsequently purchase something from the advertised Web site. Online merchants such as eBay and Dell are among the larger companies which use affiliate marketing. In order for affiliate marketing to work, the affiliate places a tag such as a cookie or a session variable on the user's request, which the merchant associates with any purchases made. The affiliate then receives a small commission.

Spyware which attacks affiliate networks does so by placing the spyware operator's affiliate tag on the user's activity—replacing any other tag, if there is one. This harms just about everyone involved in the transaction other than the spyware operator. The user is harmed by having their choices thwarted. A legitimate affiliate is harmed by having their earned income redirected to the spyware operator. Affiliate marketing networks are

harmed by the degradation of their reputation. Vendors are harmed by having to pay out affiliate revenues to an "affiliate" who did not earn them according to contract.

Affiliate fraud is a violation of the terms of service of most affiliate marketing networks. As a result, spyware operators such as WhenU and 180 Solutions have been terminated from affiliate networks including LinkShare and ShareSale.

Identity theft and fraud

One case has closely associated spyware with identity theft. In August 2005, researchers from security software firm Sunbelt Software believed that the makers of the common CoolWebSearch spyware had used it to transmit "chat sessions, user names, passwords, bank information, etc.", but it turned out that "it actually is its own sophisticated criminal little trojan that's independent of CWS." This case is currently under investigation by the FBI.

Spyware has pricipally become associated with identity theft in that keyloggers get routinely packaged within spyware. John Bambenek, who researches information security, estimates that identity-thieves have stolen over $24 billion US dollars worth of account information in the United States alone .

Spyware-makers may perpetrate another sort of fraud with *dialer* program spyware: wire fraud. Dialers cause a computer with a modem to dial up a long-distance telephone number instead of the usual ISP. Connecting to the number in question involves long-distance or overseas charges, this can result in massive telephone bills, which the user must either pay or contest with the telephone company. Dialers are somewhat less effective today, now that fewer Internet users use dialup modems.

Digital rights management

Some copy-protection schemes, while they do serve the purpose of attempting to prevent piracy, also behave similarly to spyware programs. Some digital rights management technologies (such as Sony's XCP) actually use trojan-horse tactics to verify a user as the rightful owner of the media in question.

Spyware and cookies

Anti-spyware programs often report Web advertisers' HTTP cookies as spyware. Web sites (including advertisers) set cookies — small pieces of data rather than software—to track Web-browsing activity: for instance to maintain a "shopping cart" for an online store or to maintain consistent user settings on a search engine.

Only the Web site that sets a cookie can access it. In the case of cookies associated with advertisements, the user generally does not intend to visit the Web site which sets the cookies, but gets redirected to a cookie-setting third-party site referenced by a banner ad image. Some Web browsers and privacy tools offer to reject cookies from sites other than the one that the user requested.

Advertisers use cookies to track people's browsing among various sites carrying ads from the same firm and thus to build up a marketing profile of the person or family using the

computer. For this reason many users object to such cookies, and anti-spyware programs offer to remove them.

User consent and legality

Gaining unauthorized access to a computer is illegal, under computer crime laws such as the United States Computer Fraud and Abuse Act. Since the owners of computers infected with spyware generally claim that they never authorized the installation, a *prima facie* reading would suggest that the promulgation of spyware would count as a criminal act. Law enforcement has often pursued the authors of other malware programs, such as viruses. Nonetheless, few prosecutions of writers of spyware have occurred, and many such producers operate openly as aboveboard businesses. Some have, however, faced lawsuits.

Spyware producers primarily argue in defense of the legality of their acts that, contrary to the users' claims, users do in fact give consent to the installation of their spyware. Spyware that comes bundled with shareware applications may appear, for instance, described in the legalese text of an end-user license agreement (EULA). Many users habitually ignore these purported contracts, but spyware companies such as Claria claim that these demonstrate that users have consented to the installation of their software.

Despite the ubiquity of EULAs and of clickwrap agreements, relatively little case law has resulted from their use. It has been established in most common law jurisdictions that a clickwrap agreements can be a binding contract in certain circumstances. This does not however mean that every clickwrap agreement is a contract or that every term in a clickwrap contract is enforceable. It seems highly likely that many of the purported contract terms presented in clickwrap agreements would be dismissed in most jurisdictions as being contrary to public policy. Many spyware clickwrap agreements appear intentionally ambiguous and excessive in length, with key contract terms made inconspicuous. These are all grounds on which similar agreements have been rejected as contracts of adhesion.

Nor can a contract possibly exist in the case of spyware installed by surreptitious means, such as in a drive-by download where the user receives no opportunity to either agree to or refuse the contract terms.

Some jurisdictions, including the U.S. states of Iowa and Washington, have passed laws criminalizing some forms of spyware. Such laws make it illegal for anyone other than the owner or operator of a computer to install software that alters Web-browser settings, monitors keystrokes, or disables computer-security software.

New York Attorney General Eliot Spitzer has pursued spyware companies for fraudulent installation of software. [9] In a suit brought in 2005 by Spitzer, the California firm Intermix Media, Inc. ended up settling by agreeing to pay $7.5 million and to stop distributing spyware. Intermix's spyware spread via drive-by download, and deliberately installed itself in ways that made it difficult to remove.

Another spyware behavior has attracted lawsuits: the replacement of Web advertisements. In June 2002, a number of large Web publishers sued Claria for replacing advertisements, but settled out of court. Other spyware apart from Claria's also replaces advertisements, thus diverting revenue from the ad-bearing Web site to the spyware author.

One legal issue not yet pursued involves whether courts can hold advertisers responsible for spyware which displays their ads. In many cases, the companies whose advertisements appear in spyware pop-ups do not directly do business with the spyware firm. Rather, the advertised company contracts with an advertising agency, which in turn contracts with an online subcontractor who gets paid by the number of "impressions" or appearances of the advertisement. Some major firms such as Dell Computer and Mercedes-Benz have "fired" advertising agencies which have run their ads in spyware.

In a sort of turnabout, a few spyware companies have threatened websites which have posted descriptions of their products. In 2003, Gator (now known as Claria) filed suit against the website PC Pitstop for describing the Gator program as "spyware". PC Pitstop settled, agreeing not to use the word "spyware", but continues to publish descriptions of the harmful behavior of the Gator/Claria software.

Remedies and prevention

As the spyware threat has worsened, a number of techniques have emerged to counteract it. These include programs designed to remove or to block spyware, as well as various user practices which reduce the chance of getting spyware on a system.

Nonetheless, spyware remains a costly problem. When a large number of pieces of spyware have infected a Windows computer, the only remedy may involve backing up user data, and fully reinstalling the operating system.

Virtual Machines

Using a virtual machine (such as a pre-built Browser Appliance for VMWare Player) can inhibit infection by spyware, malware, and viruses. Virtual machines provide seperate environments, so if spyware enters the virtual environment, the host computer remains unaffected. One can also use snapshots to remove one's private information, transporting the snapshot of the VM.

This environment resembles a sandbox. It has drawbacks in that it uses more memory (compared to a standalone browser) and it uses a lot of disk space.

Security practices

To deter spyware, computer users have found a number of techniques useful in addition to installing anti-spyware software.

Many system operators install a web browser other than Microsoft's Internet Explorer (IE), such as Opera or Mozilla Firefox - though such web browsers have also suffered from some security vulnerabilities. Not a single browser ranks as safe, because in the case of spyware the security comes with the person who uses the browser.

Some Internet Service Providers — particularly colleges and universities — have taken a different approach to blocking spyware: they use their network firewalls and web proxies to block access to Web sites known to install spyware. On March 31, 2005, Cornell University's Information Technology department released a report detailing the behavior of one particular piece of proxy-based spyware, *Marketscore*, and the steps the university took to intercept it. Many other educational institutions have taken similar steps against

Marketscore and other spyware. Spyware programs which redirect network traffic cause greater technical-support problems than programs which merely display ads or monitor users' behavior, and so may attract institutional attention more readily.

Spyware may get installed via certain shareware programs offered for download. Downloading programs only from reputable sources can provide some protection from this source of attack.

Adware

Adware or **advertising-supported software** is any software package which automatically plays, displays, or downloads advertising material to a computer after the software is installed on it or while the application is being used.

Application

Adware is software integrated into or bundled with a program. It is usually seen by the programmer as a way to recover programming development costs, and in some cases it may allow the program to be provided to the user free of charge or at a reduced price. The advertising income may allow or motivate the programmer to continue to write, maintain and upgrade the software product.

Some adware is also shareware, and so the word may be used as term of distinction to differentiate between types of shareware software. What differentiates adware from other shareware is that it is primarily advertising-supported. Users may also be given the option to pay for a "registered" or "licensed" copy to do away with the advertisements.

Controversy

There are concerns about adware because it often takes the form of spyware, in which information about the user's activity is tracked, reported, and often re-sold, often without the knowledge or consent of the user. Of even greater concern is malware, which may interfere with the function of other software applications, in order to force users to visit a particular web site.

It is not uncommon for people to confuse "adware" with "spyware" and "malware", especially since these concepts overlap. For example, if one user installs "adware" on a computer, and consents to a tracking feature, the "adware" becomes "spyware" when another user visits that computer, and interacts with and is tracked by the "adware" without their consent.

Spyware has prompted an outcry from computer security and privacy advocates, including the Electronic Privacy Information Center. Often, spyware applications send the user's browsing habits to an adserving company, which then targets adverts at the user based on their interests. Kazaa and eXeem are popular programs which incorporate software of this type.

Adware programs other than spyware do not invisibly collect and upload this activity record or personal information when the user of the computer has not expected or

approved of the transfer, but some vendors of adware maintain that their application which does this is not also spyware, due to disclosure of program activities: for example, a product vendor may indicate that since somewhere in the product's Terms of Use there is a clause that third-party software will be included that may collect and may report on computer use, that this Terms of Use disclosure means the product is just adware.

A number of software applications are available to help computer users search for and modify adware programs to block the presentation of advertisements and to remove spyware modules. To avoid a backlash, as with the advertising industry in general, creators of adware must balance their attempts to generate revenue with users' desire to be left alone.

Online marketing

Online Marketing is marketing on the Internet. It is a type of e-marketing, which in turn is a type of e-commerce. While at first the confusion of experiments, beta versions of websites, search engines and other online devices cause marketers to consider this world of the Internet unknowable and perhaps too unpredictable, there is now a growing body of work to which marketers are now paying attention in order to develop online marketing programs. The most known tools to marketers in the mid 2000s are currently tools grouped into 2 fields: online advertising and search engine optimization. E -marketing tools used to drive visitors to a web site include:

However, marketing online is simply not offline marketing applied to a new online world. Online marketing has a slightly different character and purpose as indicated in such seminal works as The cluetrain manifesto, Purple cow, Permission marketing, and other texts of smaller nature compiled in blogs and news sites.

Purpose of Online Marketing

When marketing online, the general four step process of marketing is still the guiding idea, in the online world the character of marketing becomes more deeply a conversation between a marketer and a market-of-one a concept that is central to The cluetrain manifesto. In such a role as a communicator, the online marketer is in a position to build awareness of her/his company or business in more personal terms than otherwise, and in so doing enables a more human conversation. Such conversations tend to be more warts and all and should establish confidence of the potential purchaser in the potential vendor.

Smith and Chaffey (2001) claim that Internet technology can be used to focus marketing on the customer, while at the same time linking to other business operations so as to achieve profitability. This can be done by:

- **Identifying** - the Internet be used for marketing research to find out customers' needs and wants;
- **Anticipating** - the Internet provides an additional channel by which customers can access information and make purchases - understanding this demand is key to governing resource allocation to e-marketing.
- **Satisfying** - a key success factor in e-marketing is achieving customer satisfaction through the electronic channel, this raises issues such as is the site easy to use, does it perform adequately, what is the standard of associated customer service and how are physical products dispatched?

Detractors of this concept of human-to-human contact through online conversations suggest that companies are going to be careful about marketing in this manner and perhaps will never really have honest and open conversations as the interests of companies and businesses are not the interests of potential purchasers. The cluetrain manifesto allows for this type of thinking suggesting that businesses when marketing in this manner need to be

thinking about more than just making money; if a business is thinking only about making money, it will become apparent in close online conversations and the market will treat that business in whatever manner it may as markets can now talk to each other through the same means marketers talk to potential customers.

Online Marketing Activities

Smith and Chaffey (2001) describe five key online marketing activities (the '5Ss') which can be applied by an organisation to implement various online marketing tactics. For example, for an e-newsletter, the 5Ss are:

- **Sell** - Grow sales (the e-newsletter often acts as both a customer acquisition tool and a retention tool - the lastminute.com e-newsletter has this dual role)
- **Serve** - Add value (give customers extra benefits online such as an online exclusive offer or more in-depth information about your products or the industry sector)
- **Speak** - Get closer to customers by creating a dialogue, asking questions through online research surveys and learning about customers' preferences through tracking - which content are people most interested in.
- **Save** - Save costs (of print and post if you have a traditional offline e-newsletter can you reduce print runs or extend it to those customers you can't afford to communicate with)
- **Sizzle** - Extend the brand online. A newsletter keeps the brand 'front-of-mind' and helps reinforce brand values. Added value can also be delivered by the e-newsletter by informing and entertaining customers.

Capturing attention of potential customers can be as simple as advertising using some of the new advertising tools the online world provides, such as advertising on search engines, but it can also be about configuring more remarkable methods that tend to spread across many sites and capturing the imagination of many people in the process. There are at least three major configurations of links and tools that have been used to capture attention online: funnel building, buzz marketing and cool tools.

Building a sales funnel requires working with search engine optimization, email newsletter distribution, discussion board entries, advertisements, affiliate activities and more. In fact, any way that additional links can be provided so that a potential customer can begin a conversation with a business, is educated about that business' products/services, or is provided with concepts and propositions that will eventually lead to a sale. A funnel is usually laid down over time and is the result of continuous activity of marketers in online activities.

Buzz marketing tends to be a much quicker process and tends to involve less activity on behalf of marketers and requires attention of people online to spread by word-of-mouth, word-from-keyboards, to be fascinated or intrigued. Purple cow was sold largely through buzz marketing that spread by blogs relatively quickly.

Another tactic of gaining attention online is through the development and release of a cool tool. A cool tool is something that captures the imagination of the online browsing public

and it is thought to be so cool that it should be shared with online friends. This could be a video clip, standalone software that is cute such as a cartoon character that lives on a users screen, or some other device that is used often for a specific purpose, such as 3Ms Post-it Notes.

Right in the middle of a new marketing practice is eBay with its datafeed marketing. Essentially a store owner sets up his/her data in eBay and then by way of feeds make this data available to advertising avenues, such as Froogle, Yahoo Product Search and about another twenty of thirty other sites that take datafeeds. All the advertising feed services point the prospective purchaser to the eBay auction. This is perhaps a little like building a sales funnel as described above, however, it uses a specific technology that enables ease of use.

Marketing on the internet requires that one be found using keyword searches or some form of online advertising. In any case the trick to being successful in Online Marketing is being found within the top 30 search results. There are 3 ways that one can be found. 1.) natural search engine ranking (70% of searchers will skip over sponsored results and start with the naturally ranked sites) 2.) Paid inclusion and 3.) Pay per click. Due to the extreme difficulty of achieving a natural high ranking on a major search engine most companies opt for #'s 2 and 3 for their online marketing. Unfortunately the 3rd option is very costly and only the most well heeled companies can afford to market online via pay per click.

What is true of Online Marketing today is that one must pay to play. Since the dot com bust several years ago search engines have discerned that in order to survive and thrive they must generate significant revenue. At first the hope was that banner advertising would be sufficient to fill the search engine coffers but it soon became evident that searchers did not respond to banners. It then became evident that there were 2 primary ways to create income for search engines and online directories. Thus paid inclusion and pay per click were born!

Recently potential greed-related challenges have emerged. There are companies that create false hits and traffic. Most recently Google has been sued for click fraud. Whether or not the charges prove to be true, actions like this make people think twice about using pay per click as part of their online marketing package.

Semantic logic will allow searchers to use not just keywords to search, but rather they will search using common language. This is a big departure from the crude Boolean logic which has served the Internet searching community for the last decade.

Internet marketing

Internet marketing is the use of the Internet to advertise and sell goods and services. Internet Marketing includes pay per click advertising, banner ads, e-mail marketing, search engine marketing (including search engine optimization), blog marketing, and article marketing.

Definition and Scope

Internet marketing is a component of electronic commerce. Internet marketing can include information management, public relations, customer service, and sales. Electronic commerce and Internet marketing have become popular as Internet access is becoming more widely available and used. Well over one third of consumers who have Internet access in their homes report using the Internet to make purchases.

History

Internet marketing first began in the early 1990s as simple, text-based websites that offered product information. It then evolved into advertisements complete with graphics. The most recent step in this evolution was the creation of complete online businesses that use the Internet to promote and sell their services and goods.

Business Models and Formats

Internet marketing is associated with several business models. The main models include business-to-business and business-to-consumer (B2C). B2B consists of companies doing business with each other, whereas B2C involves selling directly to the end consumer. When Internet marketing first began, the B2C model was first to emerge. B2B transactions were more complex and came about later. A third, less common business model is peer-to-peer (P2P), where individuals exchange goods between themselves. An example of P2P is Napster, which is built upon individuals sharing files.

Internet marketing can also be seen in various formats. One version is name-your-price (e.g. Priceline.com). With this format, customers are able to state what price range they wish to spend and then select from items at that price range. With find-the-best-price websites (e.g. Hotwire.com), Internet users can search for the lowest prices on items. A final format is online auctions (e.g. Ebay.com) where buyers bid on listed items.

Benefits

Some of the benefits associated with Internet marketing include the availability of information. Consumers can log onto the Internet and learn about products, as well as purchase them, at any hour. Companies that use Internet marketing can also save money because of a reduced need for a sales force. Overall, Internet marketing can help expand from a local market to both national and international marketplaces.

Limitations

Limitations of Internet marketing create problems for both companies and consumers. Slow Internet connections can cause difficulties. If companies put too much information on their website, Internet users may struggle to load the web page. Also, Internet marketing does not allow shoppers to touch or try-on items before purchasing them.

Security Concerns

For both companies and consumers that participate in online business, security concerns are very important. Many consumers are hesitant to buy items over the Internet because they do not trust that their personal information will remain private. Recently, some

companies that do business online have been caught giving away or selling information about their customers. Several of these companies have guarantees on their websites, claiming customer information will be private. By selling customer information, these companies are breaking their own, publicized policy. Some companies that buy customer information offer the option for individuals to have their information removed from the database (known as opting out). However, many customers are unaware that their information is being shared and are unable to stop the transfer of their information between companies.

Security concerns are of great importance and online companies have been working hard to create solutions. Encryption is one of the main methods for dealing with privacy and security concerns on the Internet. Encryption is defined as the conversion of data into a form called a cipher. This cipher cannot be easily intercepted unless an individual is authorized by the program or company that completed the encryption. In general, the stronger the cipher, the better protected the data is. However, the stronger the cipher, the more expensive encryption becomes.

Effects on Industries

Internet marketing has had a large impact on several industries including music, banking, and flea markets. In the music industry, many consumers have begun buying and downloading MP3s over the Internet instead of simply buying CDs. The debate over the legality of downloading MP3s has become a major concern for those in the music industry.

Internet marketing has also affected the banking industry. More and more banks are offering the ability to perform banking tasks online. Online banking is believed to appeal to customers because it is more convenient then visiting bank branches. Currently, over 50 million U.S. adults now bank online. Online banking is now the fastest-growing Internet activity. The increasing speed of Internet connections is the main reason for the fast-growth. Of those individuals who use the Internet, 44% now perform banking activities over the Internet.

As Internet auctions have gained popularity, flea markets are struggling. Unique items that could previously be found at flea markets are being sold on Ebay.com instead. Ebay.com has also affected the prices in the industry. Buyers and sellers often look at prices on the website before going to flea markets and the Ebay.com price often becomes what the item is sold for. More and more flea market sellers are putting their items up for sale online and running their business out of their homes.

e-Marketing

E-marketing is a type of marketing that can be defined as achieving objectives through the use of electronic communications technology such as Internet, e-mail, Ebooks, database, and mobile phone. It is a more general term than online marketing which is limited to the use of internet technology to attain marketing objectives.

Dave Chaffey, working from a relationship marketing perspective, has defined e-marketing as:

Applying Digital technologies which form online channels (Web, e-mail, databases, plus mobile/wireless & digital TV) to contribute to marketing activities aimed at achieving profitable acquisition and retention of customers (within a multi-channel buying process and customer lifecycle) through improving our customer knowledge (of their profiles, behaviour, value and loyalty drivers), then delivering integrated targeted communications and online services that match their individual needs.

Chaffey's definition emphasises that:

- It should not be the technology that drives e-marketing, but the business returns from gaining new customers and maintaining relationships with existing customers.
- It also emphasises how e-marketing does not occur in isolation, but is most effective when it is integrated with other communications channels such as telemarketing, direct-mail, personal selling, advertising, publicity, sales promotion, and other promotional techniques.
- Online channels should also be used to support the whole buying process from pre-sale to sale to post-sale and further development of customer relationships where this is appropriate.
- It should be based on knowledge of customer needs developed by researching their characteristics, behaviour, what they value, and what keeps them loyal.
- The web and e-mail communications should be personally tailored to individual buyers based on the information obtained in the research.

Promotion

Promotion is one of the four aspects of marketing. The other three parts of the marketing mix are product management, pricing, and distribution. Promotion involves disseminating information about a product, product line, brand, or company.

Promotion comprises four subcategories:

- Advertising
- Personal selling
- Sales promotion
- Publicity and public relations

The specification of these four variables creates a promotional mix or promotional plan. A promotional mix specifies how much attention to pay to each of the four subcategories, and how much money to budget for each. A promotional plan can have a wide range of objectives, including: sales increases, new product acceptance, creation of brand equity, positioning, competitive retaliations, or creation of a corporate image.

Example

The publicity for the 40th anniversary of the 1966 NCAA Basketball championship included

- The renaming of a city street
- A tie-in with an autobiography with the same title
- The screening of a film with the same title
- The release of a breakfast cereal box with coordinated materials
- A pep rally on a university campus
- Media coverage

Example 2: Veranda Park

A new residential development, Veranda Park, while under construction in Orlando, Florida, USA was promoted on-site using adverisements on the construction-site fence known as Fence Mesh. A brief look at how the marketing team met the basic marketing objectives:

1. Sales Increases

Prospective customers driving or walking in proximity of the outer perimeter of the new development were made aware of the real estate available for purchase.

2. New Product Acceptance

The public was made aware of the new construction a long time before it was finished. This gave the public a chance to gradually accept the new look of this area of the neighborhood. Rather than looking at piles of dirt and construction equipment, the fence mesh treated the public to a much more thoughtful view.

3. Brand Equity

The fence mesh built brand equity by prominently portraying the development's name in a pleasing and artistic fashion. Compare this development to one without such a fence advertisement. Prospective buyers are much more likely to remember a development with fence mesh advertising over one which does not.

4. Positioning

The fence created an extremely distinct image of the high quality type of development under construction by use of images and text describing the available services and types of real estate.

5. Corporate Image

It would be fair to say that a corporation which cared enough about the appearance of their construction site to design, install and maintain a tasteful and artistic representation of their finished product on such a large scale might be perceived in a positive light.

Branding

In marketing, a **brand** is the symbolic embodiment of all the information connected with a company, product or service. A brand typically includes a name, logo, and other visual elements such as images, fonts, color schemes, or symbols. It also encompasses the set of expectations associated with a product or service which typically arise in the minds of people. Such people include employees of the brand owner, people involved with distribution, sale or supply of the product or service, and ultimately consumers.

In other contexts the term "brand" may be used where the legal term trademark is more appropriate.

Concepts

Some marketers distinguish the psychological aspect of a brand from the experiential aspect. The experiential aspect consists of the sum of all points of contact with the brand and is known as the **brand experience**. The psychological aspect, sometimes referred to as the **brand image**, is a symbolic construct created within the minds of people and consists of all the information and expectations associated with a product or service. The unicist approach to brand building considers the conceptual structure of brands, businesses and people.

Marketers seek to develop or align the expectations comprising the brand experience through **branding**, so that a brand carries the "promise" that a product or service has a certain quality or characteristic which make it special or unique. A brand image may be developed by attributing a "personality" to or associating an "image" with a product or service, whereby the personality or image is "branded" into the consciousness of consumers. A brand is therefore one of the most valuable elements in an advertising theme, as it demonstrates what the brand owner is able to offer in the marketplace. The art of creating and maintaining a brand is called brand management. You're creating the story.

A brand which is widely known in the marketplace acquires **brand recognition**. Where brand recognition builds up to a point where a brand enjoys a mass of positive sentiment in the marketplace, it is said to have achieved **brand franchise**. One goal in brand recognition is the identification of a brand without the name of the company present. Disney has been successful at branding with their particular script font (originally Walt Disney's signature, but later translated to go.com).

Brand equity measures the total value of the brand to the brand owner, and reflects the extent of brand franchise. The term **brand name** is often used interchangeably with "brand", although it is more correctly used to specifically denote written or spoken linguistic elements of a brand. In this context a "brand name" constitutes a type of trademark, if the brand name exclusively identifies the brand owner as the commercial source of products or services. A brand owner may seek to protect proprietary rights in relation to a brand name through trademark registration.

The act of associating a product or service with a brand has become part of pop culture. Most products have some kind of brand identity, from common table salt to designer clothes. In non-commercial contexts, the marketing of entities which supply ideas or

promises rather than product and services (eg. political parties or religious organizations) may also be known as "branding".

Consumers may look on branding as an important value added aspect of products or services, as it often serves to denote a certain attractive quality or characteristic. From the perspective of brand owners, branded products or services also command higher prices. Where two products resemble each other, but one of the products has no associated branding (such as a generic, store-branded product), people may often select the more expensive branded product on the basis of the quality of the brand or the reputation of the brand owner.

Advertising spokespersons have also become part of some brands, for example: Mr. Whipple of Charmin toilet tissue and Tony the Tiger of Kellogg's.

History

Brands in the field of marketing originated in the 19th century with the advent of packaged goods. Industrialization moved the production of many household items, such as soap, from local communities to centralized factories. When shipping their items, the factories would literally brand their logo or insignia on the barrels used, which is where the term comes from.

These factories, generating mass-produced goods, needed to sell their products to a wider market, to a customer base familiar only with local goods. It quickly became apparent that a generic package of soap had difficulty competing with familiar, local products. The packaged goods manufacturers needed to convince the market that the public could place just as much trust in the non-local product.

Around 1900, James Walter Thompson published a house ad explaining trademark advertising. This was an early commercial explanation of what we now know as branding.

Many brands of that era, such as Uncle Ben's rice and Kellogg's breakfast cereal furnish illustrations of the problem. The manufacturers wanted their products to appear and feel as familiar as the local farmers' produce. From there, with the help of advertising, manufacturers quickly learned to associate other kinds of brand values, such as youthfulness, fun or luxury, with their products. This kickstarted the practice we now know as branding.

Modern branding practices are studied and analyzed at research institutes such as the Zyman Institute of Brand Science at the Goizueta Business School at Emory University.

Publicity

Publicity is the deliberate attempt to manage the public's perception of a product or organization. The product could include anything from traditional goods and services, to celebrities, or works of entertainment.

From a marketing perspective, publicity is one of the variables that comprise the promotional mix. The other components of promotions are advertising, sales promotion, and personal selling. promotion is one of the variables that comprise the marketing mix.

Publicity is a tool of public relations. Whereas public relations is the management of all communication between the firm and selected target audiences, publicity is the management of product or brand related communications between the firm and the general public. It is primarily an informative activity (as opposed to a persuasive one), but its ultimate goal is to promote the organization's products, services, or brands. A publicity plan is a planned program aimed at obtaining favorable media coverage for an organization's products - or for the organization itself, to enhance its reputation and relationships with stakeholders.

A basic tool of the publicist is the press release, but other techniques include telephone press conferences, in-studio media tours, multi-component video news releases (VNR's), newswire stories, and internet releases. For these releases to be used by the media, they must be of interest to the public (or at least to the market segment that the media outlet is targeted to). The releases are often customized to match the media vehicle that they are being sent to. Getting noticed by the press is all about saying the right thing at the right time. A publicist is continuously asking what about you or your company will pique the reader's curiosity and make a good story? The most successful publicity releases are related to topics of current interest. These are referred to as news pegs. An example is if three people die of water poisoning, an alert publicist would release stories about the technology embodied in a water purification product.

But the publicist cannot wait around for the news to present opportunities. They must also try to create their own news. Examples of this include:

- Contests
- Walkathons
- Art exhibitions
- Event sponsorship
- Arrange a speech or talk
- Make an analysis or prediction
- Conduct a poll or survey
- Issue a report
- Take a stand on a controversial subject
- Arrange for a testimonial
- Announce an appointment
- Celebrate an anniversary
- Invent then present an award
- Stage a debate

- Organize a tour of your business or projects
- Issue a commendation

The advantages of publicity are low cost, and credibility (particularly if the publicity is aired in between news stories like on evening TV news casts). New technologies such as weblogs, web cameras, web affiliates, and convergence (phone-camera posting of pictures and videos to websites) are changing the cost-structure. The disadvantages are lack of control over how your releases will be used, and frustration over the low percentage of releases that are taken up by the media.

Publicity draws on several key themes including birth, love, and death. These are of particular interest because they are themes in human lives which feature heavily throughout life. In television serials several couples have emerged during crucial ratings and important publicity times, as a way to make constant headlines. Also known as a publicity stunt, the pairings may or may not be truthful.

Publicists

A publicist is a person whose job is to generate and manage publicity for a product, public figure, especially a celebrity, or for a work such as a book or movie. Publicists usually work at large companies handling multiple clients.

Effectiveness of Publicity

The theory *any press is good press* has been coined to describe situations where bad behaviour by people involved with an organisation or brand has actually resulted in positive results, due to the fame and press coverage accrued by such events.

A good example would be Paris Hilton's many antics, from lurid sex tapes to clumsy behaviour on TV shows actually increasing business at the family's chain of Hilton Hotels.

Another example would be the Australian Tourism Board's "So where the Bloody Hell are you?" Advertising Campaign that was initially banned in the UK, but the amount of publicity this generated resulted in the official website for the campaign being swamped with requests to see the banned ad.

Definition: ***Publicity*** is the means of using an external entity (celebrities, people from the media, etc) to increase the awareness levels of the product, company, goods etc amongst the public and/or buying segment.

Search engine marketing

In Internet marketing, **search engine marketing**, or **SEM**, is a set of marketing methods to increase the visibility of a website in search engine results pages (SERPs). The three main methods are:

- **Search engine optimization**, or improving rankings for relevant keywords in search results by rectifying the website structure, and content such that they could be easly read and understood by the search engine's software programs. It is seen

that website containing the latest trends and updates are first available to the visitor.

- **Search engine advertising**, or paying the search engine company for a guaranteed high ranking or an ad displayed aside the results (commonly known as pay per click advertising).
- **Paid inclusion**, or paying the search engine company for a guarantee that the website is included in their natural search index.

Search engine marketers are experts and firms who explore of weaknesses and strengths in the methods and individual products to find the best way to promote a particular website in search engines.

Methods

Search engine optimization

Search engine optimization aims to index and improve rankings for the webpages which are most relevant to the keywords searched for according to the algorithm of each search engine. The relevant pages are returned in search engine result pages(SERPS). Basically this is done by writing a naural copy of each page containing the keywords that genuinely represent the goods and the services described within the corresponding webpages. Keywords are also used in the Title Pages, Meta Tags, Headings within a density of about 6% i.e., about 6 keywords spread over a page containing 100 words.

In order to further fine tune the pages and keep them user and search engine friendly, the architecture of the website, including its internal link structure, navigation etc., are also suitably modified for human beings and search spiders to nevigate through whole wbsite pages. Search spiders then can scan all necessary data about the whole site and store in engines' data base. A good navigation systems imparts excellent experience to the users and they tend to visit the site again and again. This a sign of good achievement.

Numbers of inbound links to the site and the 'quality' of the links determine the Reputation of the website within the industry it belong to. This Reputation is one of the most important criteria for search engines to consider higher levels of rank to the deserving webpages. Algorithms are evolutionary and strives to develop every day in an attempt to provide most relevant & useful pages to the users and strike out the websites that trick them and attain higher positions for a while.

These processes are known as Organic or Algorithmic Search Engine Optimization (SEO) of websites. Eventually it is essential for each and every website to get optimized organically, though temprarity they can make use of Pay-per-Click (PPC) to market their website without having to wait for the results of Organic SEO. However, users still prefer Organic Result Pages than Pages for which Advertising charges are paid to search engines. So far for inclusion in Organic Result Pages no fees are prescribed except the high usefulness of the information to the users.

Search engine advertising

Advertising with search engines is known by different names. It is also called sponsored search. Advertising with search engines could be further classified as follows:

Advertising based on a keyword search

- Advertising based on a keyword search could take place through a search engine such as google.com, or a search engine partner site, such as shopping.com. For example, Google offers a service called AdWords, which allows companies, for a small fee, to have a link to their website featured when a user searches a specific keyword which the company specified.

Advertising based on content context

- Many search engines (e.g. Google, Yahoo! Search) have partner websites with specific content. The websites agree to let the search engines place content-specific advertising on their website, in return for a fee. The search engine then finds companies interested in advertising on websites with their desired content. For example, an online dog food retailer might have their advertisement placed on a site about dogs.

Both of these advertising formats allow advertisers to target specific users with certain interests. Generally these advertisements are paid for based on either a pay per click campaign or an impression based campaign.

Paid inclusion

Search engines use computer programs called spiders or web crawlers to automatically discover websites and catalog their content. As this process can take some time and requires a website to be linked to from another website (to allow the crawler to find it), most search engines except for Google provide another channel to be included in search rankings via paying. This is different from pay per click advertising because the inclusion is guaranteed but not placement.

Ethical considerations

Many forms of search engine optimization only amount to ensuring compliance to search engines' guidelines for inclusion and removing any technical barriers that might keep the website from reaching a proper ranking. However, other methods of search engine optimization such as keyword spamming are often viewed as "gaming the system" and considered unethical.

Displaying advertisements or sponsored results in an area visually separated from the algorithmically determined results is generally considered ethical. However, some search engines allow the ranking of a website to be influenced with a payment and provide little or no indication to the end-user that this has happened. Since the search engines give the impression or claim that the rankings reflect the relevance or popularity of the websites, this is often seen unfair or deceptive.

Search engine advertising products that don't guarantee a specific ranking or an amount of visibility are seen as unethical by some search engine marketers. The product might

provide an unspecified "boost" or the final ranking or visibility might be a result of an auction.

Paid inclusion has not caused much concern. However, it has been suggested that search engines should improve the speed they pick up new websites and that paid inclusion services thus create a conflict of interest that discourages improving service levels across the board.

Web traffic

Web traffic is the amount of data sent and received by visitors to a web site. It is a large portion of Internet traffic. This is determined by the number of visitors and the number of pages they visit. Sites monitor the incoming and outgoing traffic to see which parts or pages of their site are popular and if there are any apparent trends, such as one specific page being viewed mostly by people in a particular country. There are many ways to monitor this traffic and the gathered data is used to help structure sites, highlight security problems or indicate a potential lack of bandwidth – not all web traffic is welcome.

Some companies offer advertising schemes that, in return for increased web traffic (visitors), pay for screen space on the site. Sites also often aim to increase their web traffic through inclusion on search engines.

Measuring web traffic

Web traffic is measured to see the popularity of web sites and individual pages or sections within a site.

Web traffic can be analysed by viewing the traffic statistics found in the web server log file, an automatically-generated list of all the pages served. A *hit* is generated when any file is served. The page itself is considered a file, but images are also files, thus a page with 5 images could generate 6 hits (the 5 images and the page itself). A *page view* is generated when a visitor requests any page within the web site – a visitor will always generate at least one page view (the main page) but could generate many more.

Tracking applications external to the web site can record traffic by inserting a small piece of HTML code in every page of the web site.

Web traffic is also sometimes measured by packet sniffing and thus gaining random samples of traffic data from which to extrapolate information about web traffic as a whole across total Internet usage.

The following types of information are often collated when monitoring web traffic:

- The number of visitors
- The average number of page views per visitor – a high number would indicate that the average visitors go deep inside the site, possibly because they like it or find it useful. Conversely, it could indicate an inability to find desired information easily.
- Average visit duration – the total length of a user's visit

- Average page duration – how long a page is viewed for
- Domain classes – the top level domain of the ISP a visitor uses, useful for finding out geographical statistics
- Busy times – the most popular viewing time of the site would show when would be the best time to do promotional campaigns and when would be the most ideal to perform maintenance
- Most requested pages – the most popular pages
- Most requested entry pages – the entry page is the first page viewed by a visitor and shows which are the pages most attracting visitors
- Most requested exit pages – the most requested exit pages could help find bad pages, broken links or the exit pages may have a popular external link
- Top paths – a path is the sequence of pages viewed by visitors from entry to exit, with the top paths identifying the way most customers go through the site
- Referrers; The host can track the (apparent) source of the links and determine which sites are generating the most traffic for a particular page.

Web sites like Alexa Internet produce traffic rankings and statistics based on those people who access the sites while using the Alexa toolbar. The difficulty with this is that it's not looking at the complete traffic picture for a site. Large sites usually hire the services of companies like Nielsen Netratings, but their reports are available only by subscription.

Controlling web traffic

The amount of traffic seen by a web site is a measure of its popularity. By analysing the statistics of visitors it is possible to see shortcomings of the site and look to improve those areas. It is also possible to increase (or, in some cases decrease) the popularity of a site and the number of people that visit it.

Limiting access

It is sometimes important to protect some parts of a site by password, allowing only authorised people to visit particular sections or pages.

Some site administrators have chosen to block their page to specific traffic, such as by geographic location. The re-election campaign site for U.S. President George W. Bush (GeorgeWBush.com) was blocked to all internet users outside of the U.S. on 25 October 2004 after a reported attack on the site.

It is also possible to limit access to a web server both based on the number of connections and by the bandwidth expended by each connection. On Apache HTTP servers, this is accomplished by the *limitipconn* module and others.

Increasing web traffic

Web traffic can be increased by placement of a site in search engines and purchase of advertising, including bulk e-mail, pop-up ads, and in-page advertisements. Web traffic can also be increased by purchasing non-internet based advertising.

If a web page is not listed in the first pages of any search, the odds of someone finding it diminishes greatly (especially if there is other competition on the first page). Very few people go past the first page, and the percentage that go to subsequent pages is substantially lower. Consequently, getting proper placement on search engines is as important as the web site itself.

There are a number of other things you can do to increase your web traffic, including but not limited to building link popularity, offering free e-books or articles and classified advertisements.

Of the above mentioned items,perhaps the easiest one to do is building link popularity. This can be accomplished by writing e-mails to sites similar to yours and asking if they would link to your site.The second way of increasing your web traffic is writing to e-zines or to free article sites.There are many sites which will accept your written material, the catch is that you are giving it away for free. The benefit however is that you get to include a link to site in the article. Meaning everytime someone clicks on your link, it is free traffic for your.

Organic traffic

Web traffic that comes from unpaid listing at search engines or directories is commonly known as "Organic" traffic. Organic Traffic can be generated/increased by including the web site in Directories (p.e. Yahoo, DMOZ), Search Engines (p.e. Google, Inktomi), Guides (p.e. Yellow Pages, Restaurant Guides) and Award Sites.

In most cases the best way to increase web traffic is to register it with the major search engines. Just registering does not guarantee traffic, as search engines work by "crawling" registered web sites. These crawling programs (crawlers) are also known as "spiders" or "robots". Crawlers start at the registered home page, and usually follow the hyperlinks it finds, to get to pages inside the web site (internal links). Crawlers start gathering information about those pages and storing it and indexing it in the search engine database. In every case, they index the page URL and the page title. In most cases they also index the Web page header (meta tag) and a certain amount of the text of the page. Then, when a search engine user looks for a particular word or phrase, the search engine looks into the database and produces the results, usually sorted by relevance according to the search engine algorithms.

Usually, the top organic result gets most of the clicks from internet users. According to some studies the top result gets between 5% and 10% of the clicks. Each subsequent result gets between 30% and 60% of the clicks of the previous one. So it is definitely important to appear in the top results. There are some companies that specialize in search engine marketing. However, it is becoming common for webmasters to get approached by "boiler-room" companies with no real knowledge of how to get results. As opposed to Pay per Clicks, search engine marketing is usually paid monthly or yearly, and most search engine companies cannot promise specific results for what is paid to them.

Because of the huge amount of information available on the internet, crawlers might take days, weeks or months to complete review and index all the pages they find. Google, for example, as of the end of 2004 had indexed over 8 billion pages. Even having hundreds or thousands of servers working on the spidering of pages, a complete reindex takes its time.

That is why some pages recently updated in certain web sites are not immediately found when doing searches on search engines.

Paid advertising
In return for a small payment many larger companies choose to advertise their sites on other popular sites. This e-marketing usually takes the form of:

- Banner advertising: Banner impressions are sold by the thousands, and referred to as Cost Per Impression (CPM). As of 2004, prices range from $1/CPM for a run-of-network to about $50/CPM or more for specialized targeted runs. Most popular web sites sell banner advertising space, with the notable exception of Google.
- Pay per clicks: Advertisers "buy" keywords or keyphrases by bidding on them against other advertisers. The so called Pay-per-click engines sell their premium spaces showing in the searches the highest paying advertisers. Google sells paid advertisement through its AdWords and AdSense systems, which place sponsored links on search pages. Overture, now owned by Yahoo!, is one of the most popular pay-per-click advertising venues.

As users got used to seeing banners, some companies chose to make the advertisements more intrusive – pop-up ads became particularly popular to attract attention. However, most people consider pop-ups a nuisance and several software companies offer free pop-up blockers. Even Microsoft included a pop-up blocker in Service Pack 2 of Windows XP.

Traffic overload

Too much web traffic can dramatically slow down or even prevent all access to a web site. This is caused by more file requests going to the server than it can handle and may be an intentional attack on the site or simply caused by over-popularity. Large scale web sites with numerous servers can often cope with the traffic required and it is more likely that smaller services are affected by traffic overload.

Denial of service attacks

Denial-of-service attacks (DoS attacks) have forced web sites to close after a malicious attack, flooding the site with more requests than it could cope with. Viruses have also been used to co-ordinate large scale distributed denial-of-service attacks.

Sudden popularity

A sudden burst of publicity may accidentally cause a web traffic overload. A news item in the media, a quickly-propogating email, or a link from a popular site may cause such a boost in visitors (sometimes called a flash crowd) that overwhelms the site.

Web sites have been forced to close after an unexpected mass increase of traffic, particularly those run by an individual leasing the bandwidth from an ISP or hosting site. Some sites backed by large companies running their own servers have also been caught out by the problems of overpopularity. When first announced, the Vision of Britain Through Time site, containing information taken from the 1901 UK census, was advertised on numerous television programmes and causing such interest that the site had to be taken offline until different arrangements were made to cope with the traffic. The site was hosted

by a project at the University of Edinburgh and they had not foreseen the amount of bandwidth and the server load that would be required. Ironically, by the time the site was able to cope with the traffic both the interest and the free advertisements of the site had greatly slowed, giving them excess capacity.

There are some particular web sites that are so popular that any links to external sites can cause problems for the destination host.

Affiliate marketing

Affiliate Marketing is a popular method of promoting web businesses in which an affiliate is rewarded for every visitor, subscriber and/or customer provided through his efforts. It is a modern variation of the practice of paying finder's-fees for the introduction of new clients to a business. Compensation may be made based on a certain value for each visit (Pay per click), registrant (Pay per lead), or a commission for each customer or sale (Pay per Sale), or any combination.

The most attractive aspect of affiliate marketing, from the merchant's viewpoint, is that with this pay for performance model, no payment is due to an affiliate until results are realized.

Some e-commerce sites run their own affiliate programs while other e-commerce vendors use third party services provided by intermediaries to track traffic or sales that are referred from affiliates. Some businesses owe much of their growth and success to this marketing technique, although research has shown in general the increase to be approximately 15-20% of online revenue.

Some advertisers offer multi-tier affiliate programs that distribute commission into a hierarchical referral network of sign-ups and sub-affiliates. In practical terms: publisher "A" signs up the affiliate program with an advertiser and gets rewarded for the agreed activity conducted by a referred visitor. If publisher "A" attracts other publishers ("B", "C", etc.) to sign up for the same affiliate program using her sign-up code all future activities by the joining publishers "B" and "C" will result in additional, lower commission for publisher "A".

Snowballing, this system rewards a chain of hierarchical publishers who may or may not know of each others' existence, yet generate income for the higher level signup. Most affiliate programs are simply one-tier.

Merchants who are considering adding an affiliate strategy to their online sales channel should research the different technological solutions available to them. Some types of affiliate management solutions include: standalone software, hosted services, shopping carts with affiliate features, and third party affiliate networks.

In its early days many internet users held negative opinions of affiliate marketing due to the tendency of affiliates to use spam to promote the programs in which they were enrolled. As affiliate marketing has matured many affiliate merchants have refined their terms and conditions to prohibit affiliates from spamming.

Currently there is much debate around the affiliate practice of Spamdexing and many affiliates have converted from sending email spam to creating large volumes of autogenerated webpages each devoted to different niche keywords as a way of SEOing their sites with the search engines. This is sometimes referred to as spamming the search engine results. Spam is the biggest threat to organic Search Engines whose goal is to provide quality search results for keywords or phrases entered by their users. Google's algorithm update dubbed "Big Daddy" in February 2006 which was the final stage of Google's major update dubbed "Jagger" which started mid-summer 2005 specifically targeted this kind of spam with great success and enabled Google to remove a large amount of mostly computer generated duplicate content from its index.

Early days

In the early days of affiliate marketing, there was very little control over what affiliates were doing, which was abused by a large number of affiliates. Affiliates used false advertisements, trademark bidding on search engines, forced clicks to get tracking cookies set on users' computers, and Adware. Many affiliate programs were poorly managed.

This changed dramatically over the last few years for multiple reasons. Revenue generated online grew quickly. The e-commerce website, viewed as a marketing toy in the early days of the web, became an integrated part of the overall business plan and in some cases grew to a bigger business than the existing offline business. Many companies hired outside affiliate management companies to manage the affiliate program.

When Google, the most popular search engine on the Internet, introduced AdWords (pay-per-click advertising pioneered by Goto.com, then Overture.com and now Yahoo! Search Marketing) many Merchants became aware of the issue of trademark bidding by affiliates. The terms of service were quickly modified by most merchants and structures were put in place to monitor affiliate activities.

Adware

Adware is still an issue today, but affiliate marketers have taken steps to fight it. Merchants usually had no clue what adware was, what it does and how it was damaging their brand. Affiliate marketers became aware of the issue much quicker, especially because they noticed that adware often overwrites their tracking cookie and results in a decline of commissions. Affiliates who do not use adware became enraged by adware, which they felt was stealing hard earned commission from them. Adware usually has no valuable purpose or provides any useful content to the often unaware user that has the adware running on his computer. Affiliates discussed the issues in various affiliate forums such as ABestWeb and started to get organized. It became obvious that the best way to cut off adware was by discouraging merchants from advertising via adware. Merchants that did not care or even supported adware were made public by affiliates, which damaged the merchants' reputations and also hurt the merchants' general affiliate marketing efforts. Many affiliates simply "canned" the merchant or switched to a competitor's affiliate program. Eventually, affiliate networks were also forced by merchants and affiliates to take a stand and ban adware publishers from their network.

The new Web

The rise of blogging, interactive online communities and other new technologies, web sites and services based on the concepts that are now called Web 2.0 have impacted the affiliate marketing world as well. The new media allowed merchants to get closer to their affiliates and improved communication between each other. New portals like Return on Affiliates allow affiliates, merchants, and networks to interconnect easily, on a professional and a personal level.

New developments have made it harder for unscrupulous affiliates to make money. Emerging black sheep are detected and made known to the affiliate marketing community with much greater speed and efficiency.

Affiliate

An **affiliate** is a commercial entity with a relationship with a peer or a larger entity.

Broadcast networks

In a radio network or TV network, an affiliate is a radio station or TV station that agrees to carry the broadcasts of, but is not owned by, the network. Usually, the stations are still responsible for the content (such as profanity) to some extent. An affiliate is not the same as an owned and operated station, which is owned by the network such a station carries programming for.

Electronic commerce

Affiliate marketing typically refers to this Electronic commerce version of the traditional agent/referral fee sales channel concept. An **e-commerce affiliate** is a website which links back to an e-commerce site such as Amazon.com. When a reader of the website clicks on a link, they are connected to the e-tailer and if they purchase something the affiliate receives a small payment, usually a percentage of the money the customer spends. Affiliates can also be referred as publishers. The Hotel and Travel Industry uses affiliate marketing to a large extent.

Corporate structure

A corporation may be referred to as an **affiliate** of another when it is related to it but not strictly controlled by it, as with a subsidiary relationship, or when it is desired to avoid the appearance of control. This is sometimes seen with multinational companies that need to avoid restrictive laws (or negative public opinion) on foreign ownership.

Affiliate networks

An affiliate network is composed of a group of merchants and a group of affiliates. Merchants join the network and affiliates join the network in order to advertise the merchant products in exchange of a commission from the merchant. Affiliate networks present some great advantages for the merchant and the affiliate. The merchant gets potential access to a wide networks of affiliates. The affiliate does not necessarily need to

make a certain sale amount for one particular merchant but rather for the entire range of merchants before getting paid.

The affiliate also puts more trust in a network rather than a merchants independent affiliate program. The merchants pay the overall commission to the network. The network then distributes the money to each affiliate who made the sale.

Use of affiliate links

Sites made up mostly of affiliate links are usually badly regarded as they do not offer quality content. In 2005 there were active changes made by Google whereby certain websites were labeled as "thin affiliates" and were either removed from the index, or taken from the first 2 pages of the results and moved deeper within the index. In order to avoid this categorization, webmasters who are affiliate marketers must create real value within their websites that distinguishes their work from the work of spammers or banner farms with nothing but links leading to merchant sites.

Affiliate links work best in the context of the information contained within the website. For instance, if a website is about "How to publish a website", within the content an affiliate link leading to a merchant's ISP site would be appropriate. If a website is about Sports, then an affiliate link leading to a sporting goods site might work well within the content of the articles and information about sports. The idea is to publish quality information within the site, and to link "in context" to related merchant's sites.

One common use of affiliate links is shopping directories and or price comparison websites. However, these sites should do their best to enhance the web shopping experience. In many other cases, affiliate marketers offer unique content in niche subject areas that they have researched well, and their text or graphic links to a merchant's site are well placed. This principle works very well in blog website marketing as well.

AdSense

AdSense is an advertising program run by Google. Website owners can enroll in this program to enable text and image advertisements on their sites. These ads are administered by Google and generate revenue on either a per-click or per-thousand-ads-displayed basis. Google utilizes its search technology to serve ads based on website content, the user's geographical location, and other factors. Those wanting to advertise with Google's targeted ad system may sign up through AdWords. AdSense has become a popular method of placing advertising on a website because the ads are less intrusive than most banners, and the content of the ads is often relevant to the website.

It currently uses JavaScript code to incorporate the advertisements into a participating site. If it is included on a site which has not yet been crawled by the Mediabot, it will temporarily display advertisements for charitable causes known as public service announcements (PSAs). (Note that the Mediabot is a separate crawler from the Googlebot that maintains Google's search index.)

Many sites use AdSense to monetize their content and some webmasters work hard to maximize their own AdSense income. They do this in three ways:

1. They use a wide range of traffic generating techniques including but not limited to online advertising.
2. They build valuable content on their sites; content which attracts AdSense ads and which pay out the most when they get clicked.
3. They use copy on their websites that encourage clicks on Ads. Note that Google prohibits people from using phrases like "Click on my AdSense ads" to increase click rates. Phrases accepted are "Sponsored Links" and "Advertisements".

The source of all AdSense income is the AdWords program which in turn has a complex pricing model based on a Vickrey second price auction, in that it commands an advertiser to submit a sealed bid (not observable by competitors). Additionally, for any given click received, advertisers only pay one bid increment above the second-highest bid.

AdSense for feeds

In May 2005, Google unveiled **AdSense for feeds**, a version of AdSense than runs on RSS and Atom feeds that have more than 100 active subscribers. According to the Google Blog, "advertisers have their ads placed in the most appropriate feed articles; publishers are paid for their original content; readers see relevant advertising — and in the long run, more quality feeds to choose from".

AdSense for feeds works by inserting images into a feed. When the image is displayed by the reader/browser, Google writes the ad content into the image that it returns. The ad content is chosen based on the content of the feed surrounding the image. When the user clicks the image, he or she is redirected to the advertiser's site in the same way as regular AdSense ads.

AdSense for search

A companion to the regular AdSense program, **AdSense for search** lets website owners place Google search boxes on their pages. When a user searches the web or the site with the search box, Google shares any ad revenue it makes from those searches with the site owner.

Abuse of Google AdSense

Some webmasters create sites tailored to lure searchers from Google and other engines onto their AdSense to make money from clicks. These "zombie" sites often contain nothing but a large amount of interconnected, automated content (e.g. a directory with content from the Open Directory Project). Possibly the most popular form of such "AdSense farms" are splogs ("spam blogs"), which are centered around known high-paying keywords. Also many sites use the free Wikipedia content to attract visitors. These and related approaches are considered to be search engine spam and can be reported to Google.

How AdSense works

Each time a visitor visits a page with an AdSense tag, a piece of JavaScript writes an iframe tag, whose src attribute includes the URL of the page. Google's servers use a cache of the page for the URL or the keywords in the URL itself to determine a set of high-value keywords. (Some of the details are described in the AdSense patent.) If keywords have been cached already, ads are served for those keywords based on the AdWords bidding system.

The storage requirements of an AdSense system are stunningly modest. If each URL has just 8 "high-value" keywords, each represented by a single 32-bit number, then the keywords for each URL could be represented with just 32 bytes. The high value keywords of 4 billion URLs could be stored in 128GB, which would cost only $100 (circa 2006). 400 billion URLs or 100 drives (for a redundancy of 100) would require only $10,000 in storage costs.

AdSense serves a very large number of pages each day. If each day around 1B people saw 10 AdSense impressions (or 100M people saw 100 AdSense impressions), then AdSense would serve around 10B requests/day, or 115,741 requests/sec. If one machine can serve 20 reqs/second (seek times to read a random 4096-byte location on a drive allow for bursts of well over 100 reqs/second), then Google would require 5,787 servers to serve these 10B reqs/day. If each of these servers were hosted at a cost of $100/month, then it would cost $579K/month to run the adservers needed.

Suppose these 10B impressions/day generated clicks at a clickthrough rate of .3% and an average CPC of $.10. Then each day Google would receive 30M clicks/day (347 clicks/sec), generating $3M/day ($34.77/sec), or 900M clicks/month, generating $90M/month.

e-Mail marketing

E-mail marketing is a form of direct marketing which uses electronic mail as a means of communicating commercial or fundraising messages to an audience. In its broadest sense, every e-mail sent to a potential or current customer could be considered e-mail marketing. However, the term is usually used to refer to:

- Sending e-mails with the purpose of enhancing the relationship of a merchant with its current or old customers and to encourage customer loyalty and repeat business.
- Sending e-mails with the purpose of acquiring new customers or convincing old customers to buy something immediately.
- Adding advertisements in e-mails sent by other companies to their customers.

Researchers estimate that as of 2004 the E-mail Marketing industry's revenues has surpassed the $1 billion/yr mark.

The Good

E-mail marketing is popular with companies because:

- It is extremely cheap. Compared to direct mailing or printed newsletters the costs are negligible. The advertiser does not need to pay for production, paper, printing or postage.
- It is instant, as opposed to a mailed advertisement, an e-mail arrives in a few seconds or minutes.
- It lets the advertiser "push" the message to its audience, as opposed to a website that waits for customers to come in.
- It is easy to track. An advertiser can track bounce-backs, positive or negative responses, click-throughs, rise in sales.
- Advertisers can reach substantial numbers of e-mail subscribers who have opted in (consented) to receive e-mail communications on subjects of interest to them
- It has been proven successful when well done.
- When most people switch on their computer the first thing they do is check their e-mail.
- Specific types of interaction with messages can trigger other messages to be automatically delivered.

The Bad

Many companies use e-mail marketing to communicate with existing customers, but many other companies send unsolicited commercial e-mail, also known as spam.

Illicit e-mail marketing antedates legitimate e-mail marketing, since on the early Internet it was not permitted to use the medium for commercial purposes. As a result, marketers attempting to establish themselves as legitimate businesses in e-mail marketing have had an uphill battle, hampered also by criminal spam operations billing themselves as legitimate.

It is frequently difficult for observers to distinguish between legitimate and spam e-mail marketing. First off, spammers attempt to represent themselves as legitimate operators, obfuscating the issue. Second, direct-marketing political groups such as the U.S. Direct Marketing Association (DMA) have pressured legislatures to legalize activities which many Internet operators consider to be spamming, such as the sending of "opt-out" unsolicited commercial e-mail. Third, the sheer volume of spam e-mail has led some users to mistake legitimate commercial e-mail (for instance, a mailing list to which the user subscribed) for spam — especially when the two have a similar appearance, as when messages include HTML and flashy graphics.

Due to the volume of spam e-mail on the Internet, spam filters are essential to most users. Some marketers report that legitimate commercial e-mails frequently get caught by filters, and hidden; however, it is somewhat less common for e-mail users to complain that spam filters block legitimate mail.

Companies considering an e-mail marketing program must make sure that their program does not violate spam laws such as the United States' CAN-SPAM Act, the European Privacy

& Electronic Communications Regulations 2003 or their Internet provider's acceptable use policy. Even if a company follows the law, if Internet mail administrators find that it is sending spam it is likely to be listed in blacklists such as SPEWS.

E-mail marketing terms

Auto-responders

- Automatic replies sent by the e-mail software of the recipient after receipt of an e-mail.

Bounce backs

- e-mail sent back to the server that originally sent the e-mail.

Bounce rate

- Ratio of bounced e-mails to total e-mails sent.

Bulk, bulking

- Terms used by spammers to refer to their line of work. Mostly synonymous with spam or UCE.

Call to action

- Words in the e-mail that entice recipients to do something.

Click-through

- The action of clicking on a link.

Click-through rate (CTR)

- Ratio of click-throughs to total e-mails sent.

Commercial e-mail

- Any e-mail sent for commercial purpose; for instance, an advertisement to buy a product or service, an order confirmation from an online store, or a paid subscription periodical delivered by e-mail. Commercial e-mail is *not* synonymous with spam; see *unsolicited commercial e-mail* below.

Demographic

- Characteristic of a group of e-mail recipients.

Double opt-in

- A term coined by spammers to refer to the normal operation of secure electronic mailing list software. A new subscriber first gives his/her address to the list software (for instance, on a Web page) and then confirms subscription after receiving an e-mail asking if it was really him/her. This ensures that no person can subscribe someone else out of malice or error. The intention of the term "double opt-in" is to make it appear that the confirmation is a duplication of effort; and thus, to justify not confirming subscriptions. Mail system administrators and non-spam mailing list operators refer to *confirmed subscription* or *closed-loop opt-in*.

Double opt-out

- Same as Opt-In, but the recipient unsubscribes instead of subscribes. Borderline spam operations frequently make it difficult to unsubscribe from lists, in order to keep their lists large. Hard-core spam operations make it impossible -- they treat opt-out requests as confirmations that the address works and is read.

E-mail Blast

- An e-mail sent to multiple recipients, intended to inform them of announcements, events or changes. A variety of methods can be used to send the same e-mail to multiple recipients: for example: using options within an e-mail program, using the mail merge option within a word processing program, or using a commercial e-mail list programs.

Express consent

- A recipient agrees actively to subscribe by checking a box on a web form, paper form or by telephone. A recipient not unchecking a box is not express consent.

False positives

- E-mail that is not spam but is labeled spam by a spam filter of the recipient. Note that e-mail marketers may have different opinions of what is "spam" than e-mail recipients.

Format

- E-mails can be sent in plain text, HTML, or Microsoft's rich text format.

Hard bounce

- Bounced e-mail that could never get through because the e-mail address doesn't exist or the domain doesn't exist.

List broker

- Reseller of lists of e-mail addresses.

List building

- Process of generating a list of e-mail addresses for use in e-mail campaigns.

List host

- Web service that provides tools to manage large e-mail address databases and to distribute large quantities of e-mails.

List manager

- Owner or operator of opt-in e-mail newsletters or databases. Also software used to maintain a mailing list.

Look and feel

- Appearance, layout, design, functions & anything not directly related to the actual message on an e-mail.

Open rate

- E-mail open rate measures the ratio of e-mails "opened" to the number sent or "delivered." The ratio is calculated in various ways, the most popular is: e-mails delivered (sent - hard bounces) /unique opens.

Opt-in

- The action of agreeing to receive e-mails from a particular company, group of companies or associated companies, by subscribing to an e-mail list.

Opt-out

- A mailing list which transmits e-mails to people who have not subscribed and lets them "opt-out" from the list. The subscribers' e-mail addresses may be harvested from the web, USENET, or other mailing lists. ISP policies and some regions' laws consider this equivalent to spamming.

Personalization

- The use of technology and customer information to tailor e-mails between a business and each individual customer. Using information previously obtained about the customer, the e-mail is altered to fit that customer's stated needs as well as needs perceived by the business based on the available customer information, for the purpose of better serving the customer by anticipating needs, making the interaction efficient and satisfying for both parties and building a relationship that encourages the customer to return for subsequent purchases.

Privacy

- The Privacy Act of 1974, Public Law 93-579, safeguards privacy through creating four procedural rights in personal data. It requires government agencies to show an individual any records kept on him/her; also requires agencies to follow "fair information practices" when gathering and handling personal data. It places restrictions on how agencies can share an individual's data with other people and agencies and also lets individuals sue the government for violating its provisions.

Rental list

- A mailing list that can only be used once or for a limited time. The user of the list pays the owner of the list less money than if he/she would have bought the list outright. Note that this term is usually used for lists generated by address harvesting or other means; the investment made by the list creator does not correlate with the permission of the e-mail recipients. Many firms who "rent" or "buy" a list face spam complaints afterward from persons who never subscribed.

Segmentation (or Targeting)

- The use of previously gathered information to send e-mails of a particular offer to a subset of the list.

Soft bounce

- A soft bounce is an e-mail that gets as far as the recipient's mail server but is bounced back undelivered before it gets to the intended recipient. it might occur because the recipient's inbox is full. A soft bounce message may be deliverable at another time or may be forwarded manually by the network administrator in charge of redirecting mail on the recipient's domain. On the other hand, a hard bounce is an e-mail message that has been returned to the sender because the recipient's address is invalid.

Spam or UCE (Unsolicited Commercial e-mail-UCE)

- From the sender's point-of-view, spam is a form of bulk mail, often sent to a list obtained by companies that specialize in creating e-mail distribution lists. To the receiver, it usually seems like junk e-mail. Spam is equivalent to unsolicited telemarketing calls except that the user pays for part of the message since everyone shares the cost of maintaining the Internet. Spammers typically send a piece of e-mail to a distribution list in the millions, expecting that only a tiny number of readers will respond to their offer. The term spam is said to derive from a famous Monty Python sketch ("Well, we have Spam, tomato & Spam, egg & Spam, Egg, bacon & Spam...") that was current when spam first began arriving on the Internet. SPAM is a trademarked Hormel meat product that was well-known in the U.S. Armed Forces during World War II.

Spam filter

- Software that is usually installed in the users e-mail client, with the purpose of avoiding spam e-mail to get into the client's inbox or at least to be flagged as such.

Subject line

- It is one of the most important issues in e-mail marketing. The better the subject line of an e-mail, the better probability of being opened by the recipient.

Targeting (or segmentation)

- Sending e-mails to a subset of a mailing list based on a specific filter, trying to improve CTR and/or open ratios.

Tracking

- The act of reporting CTR, open ratios, bounces, etc.

Trigger based messaging

- Triggering a message based on an event or interaction with a previous message. Popular for customers who request more information

Unique click

- During a particula period, a visitor to a website could click several times on a particular link, but during that period it is counted only as one and considered a unique visitor.

Unsolicited commercial e-mail (UCE)

- Commercial e-mail, usually of an advertising nature, sent at the expense of the recipient without his or her permission. Sending UCE is an offense against all major ISPs' terms of service, and is a crime in some jurisdictions.

Opt-in e-mail advertising

Opt-in e-mail advertising or **permission marketing** is a method of advertising by electronic mail wherein the recipient of the advertisement has consented to receive it. It is one of several ways developed by marketers to eliminate the disadvantages of e-mail marketing.

E-mail has become a very popular mode of communication across the world. It has also become extremely popular to advertise through . Some of the many advantages of advertising through e-mail are the direct contact with the consumer and is "inexpensive, flexible, and simple to implement" (Fairhead, 2003). There are also disadvantages attached to e-mail advertising such as, alienating the consumer because of overload to messages or the advertisement getting deleted without getting read.

Permission e-mail marketing may evolve into a technology that uses a handshake protocol between sender and receiver (Fairhaed, 2003). This system is intended to eventually result in a high degree of satisfaction between consumers and marketers. If opt-in e-mail advertising is used, the material that is emailed to consumers will be "anticipated." It is assumed that the consumer wants to receive it, which makes it unlike unsolicited advertisements sent to the consumer (often referred to as spam). Ideally, opt-in e-mail advertisements will be more personal and relevant to the consumer than untargetted advertisements.

A common example of permission marketing is a newsletter sent to a firm's customers. Newsletters like this are a way to let customers know about upcoming events or promotions, or new products. In this type of advertising, a company that wants to send a newsletter to their customers may ask them at the point of purchase if they would like to receive this newsletter.

With a foundation of opted-in contact information stored in a database, marketers can automatically send out promotional materials. The marketers can also segment their promotions to specific market segments.

Email Marketing Services and CAN-SPAM Compliance

Because the CAN-SPAM Act of 2003 authorizes an USD 11,000 penalty per violation for spamming each individual recipient, many commercial e-mail marketers within the United States utilize a service or special software that helps ensure compliance with the Act. A variety of older systems exist which do not ensure compliance with the Act. To comply with the Act's regulation of commercial e-mail, services typically: require users to authenticate their return address and include a valid physical address, provide a one-click unsubscribe feature, and prohibit importing lists of purchased addresses which may not have given valid permission.

In addition to satisfying legal requirements, services such as ConstantContact help customers to set up and manage their own e-mail marketing campaigns. The services

provide e-mail templates, automatically handle subscriptions and removals, and generate statistics on how many messages were received and openned, and whether the recipients clicked on any links within the messages.

Permission marketing

Permission marketing is a term used in e-marketing. Marketers will ask permission before they send advertisements to prospective customers. It is used by some Internet marketers, email marketers, and telephone marketers. It requires that people first "opt-in", rather than allowing people to "opt-out" only after the advertisements have been sent.

Marketers feel that this is a more efficient use of their resources because advertisements are not sent to people that are not interested in the product. This is one technique used by marketers that have a personal marketing orientation. They feel that marketing should be done on a one-to-one basis rather than using broad aggregated concepts like market segment or target market.

The term was coined by Seth Godin in 1999 in his book of the same name.

In the United Kingdom, opt-in is required for email marketing, under The Privacy and Electronic Communications (EC Directive) Regulations 2003. This came into force on the 11 December 2003.

Telemarketing

Early History

Telemarketing is a registered trademark owned by Nadji Tehrani who founded *TeleMarketing Magazine* in 1982. Prior to that, the term was used extensively in Bell System communications relating to new uses for the outbound (WATS) and inbound (Toll-Free 800)services introduced in the late 1970s. It is a form of direct marketing where a salesperson uses the telephone to solicit prospective customers to sell products or services.

Categories

There are two major categories of telemarketing: Business-to-Business and Business-to-Consumer.

Within these two categories are two other broad divisions: Lead Generation, where the objective is to obtain information and Sales, where the object is to get someone to buy something.

Within these two categories, there are two other broad categories: Outbound and Inbound. Outbound telemarketing efforts are proactive, with the marketing person making phone calls to prospects or existing customers. Inbound telemarketing efforts are reactive, where the agent processes requests for information or takes orders. The demand is generally created by advertising, publicity or the efforts out outside salespeople.

Telemarketing may be done from a company's office, a call center or increasingly from someone's home.

Effective telemarketing programs often involve a two or more call process: The first call (or series of calls) determines the prospect or existing customer's needs. The final call (or series of calls) motivates the prospect or existing customer to make a purchase.

Negative Perceptions

The great majority of telemarketing presentations are legitimate calls from companies that offer valuable services. Unfortunately, telemarketing has also been negatively associated with various scams or frauds like multilevel marketing, pyramid schemes or with fraudulently overpriced products or services.

The prospective customers are identified and qualified by various means, including past purchase histories, previous requests for information, credit limit, competition entry forms or application forms. Names may also be purchased from another company's customer database, or obtained from a telephone directory or some other public list or forum. The qualification process is intended to find those prospective customers most likely to purchase the product or service being sold or advertised. Charitable organizations, alumni associations and political parties often use telemarketing to solicit donations.

Market survey companies often use telemarketing techniques to survey prospective or past customers of a client business to assess market acceptance or satisfaction with a particular product, service, brand or company. Public opinion polls are conducted in a similar manner.

Telemarketing techniques can also be applied to other forms of electronic marketing using e-mail or fax messages.

Telemarketing is often criticized as being an unethical business practice as some companies make unsolicited calls, using high-pressure sales techniques. Such practices may be subject to regulatory or legislative controls related to consumer privacy and protection. In particular, telemarking in the U.S. is restricted at a federal level by the FCC's Telephone Consumer Protection Act of 1991 and the FTC's Telemarketing Sales Rule. Many professional associations of telemarketers do have codes of ethics and standards that member businesses follow to win public confidence.

Do Not Call Listings

Some jurisdictions have implemented "Do Not Call" listings, either through industry organizations or legislation, in which consumers can indicate that they do not wish to be called by telemarketers. Legislative versions often provide for heavy penalties for companies calling individuals on these listings. The U.S. Federal Trade Commission has now implemented a National Do Not Call Registry in an attempt to reduce intrusive telemarketing on a national basis. Although challenged by telemarketing corporations and trade groups as a violation of commercial speech rights, the National Do Not Call Registry was upheld by the U.S. 10th Circuit Court of Appeals on February 17, 2004.

Avoiding Telemarketing Calls

There are several methods that people use to avoid telemarketing calls. Using caller ID or a privacy manager can allow the targeted subscriber to identify the caller before the call is answered and make the decision not to answer. Answering machines and voicemail can also be used to screen calls, as telemarketers generally do not leave messages. A device called the Telezapper foils telemarketing calls by issuing a tone which causes the autodialer at the call center to log the number as out of service.

Search engine optimization

Search engine optimization (SEO) is a set of methods aimed at improving the ranking of a website in search engine listings. The term also refers to an industry of consultants who carry out optimization projects on behalf of clients' sites. Practitioners may use "white hat SEO" (methods generally approved by search engines, such as building content and improving site quality), or "black hat SEO" (tricks such as cloaking and spamdexing). White hatters charge that black hat methods are an attempt to manipulate search rankings unfairly. Black hatters counter that *all* SEO is an attempt to manipulate rankings, and that the particular methods one uses to rank well are irrelevant.

Search engines display different kinds of listings in the search engine results pages (SERPs), including: pay-per-click advertisements, paid inclusion listings, and organic search results. SEO is primarily concerned with advancing the goals of a web site by improving the number and position of its organic search results for a wide variety of relevant keywords. SEO strategies can increase both the number and quality of visitors, where quality means visitors who complete the action hoped for by the site owner (e.g. purchase, sign up, learn something).

For competitive, high-volume search terms, the cost of pay per click advertising can be substantial. Ranking well in the organic search results can provide the same targeted traffic at a potentially lower cost. Site owners may choose to optimize their sites for organic search, if the cost of optimization is less than the cost of advertising.

Not all sites have identical goals for search optimization. Some sites are seeking any and all traffic, and may be optimized to rank highly for common search phrase. A broad search optimization strategy can work for a site that has broad interest, such as a periodical, a directory, or site that displays advertising with a CPM revenue model. In contrast, many businesses try to optimize their sites for large numbers of highly specific keywords that indicate readiness to buy. Overly broad search optimization can hinder marketing strategy by generating a large volume of low-quality inquiries that cost money to handle, yet result in little business. Focusing on desirable traffic generates better quality sales leads, allowing the sales force to close more business.

History

Early search engines

SEO began in the mid-1990s, as the first search engines were cataloging the early Web. Initially, all a webmaster needed to do was submit a site to the various engines which would run spiders, programs to "crawl" the site, and store the collected data. The search engines then sorted the information by topic, and served results based on pages they had spidered. As the number of documents online kept growing, and more webmasters realized the value of organic search listings, so popular search engines began to sort their listings so they could display the most relevant pages first. This was the start of a search engine versus webmaster game that continues to this day.

At first search engines were guided by the webmasters themselves. Early versions of search algorithms relied on webmaster-provided information such as category and keyword meta tags. Meta tags provided a guide to each page's content. When some webmasters began to abuse meta tags, causing their pages to rank for irrelevant searches, search engines abandoned their consideration of Meta tags and instead developed more complex ranking algorithms, taking into account factors that were more diverse, including:

- Text within the title tag
- Domain name
- URL directories and file names
- HTML tags: headings, bold and emphasized text
- Keyword density
- Keyword proximity
- Alt attributes for images
- Text within NOFRAMES tags

By relying so extensively on factors that were still within the webmasters' exclusive control, search engines continued to suffer from abuse and ranking manipulation. In order to provide better results to their users, search engines had to adapt to ensure their SERPs showed the most relevant search results, rather than useless pages stuffed with keywords by unscrupulous webmasters. This led to the rise of a new kind of search engine.

Organic search engines

Google was started by two PhD students at Stanford University, Sergey Brin and Larry Page, and brought a new concept to evaluating web pages. This concept, called PageRank, has been from the start important to the Google algorithm. PageRank relies heavily on incoming links and uses the logic that each link to a page is a vote for that page's value. The more incoming links a page had the more "worthy" it is. The value of each incoming link itself varies directly based on the PageRank of the page it comes from and inversely on the number of outgoing links on that page.

With help from PageRank, Google proved to be very good at serving relevant results. Google became the most popular and successful search engine. Because PageRank measured an off-site factor, Google felt it would be more difficult to manipulate than on-page factors.

But manipulated it was. Webmasters had already developed link manipulation tools and schemes to influence the Inktomi search engine. These methods proved to be equally applicable to Google's algorithm. Many sites focused on exchanging, buying, and selling links on a massive scale. PageRank's reliance on the link as a vote of confidence in a page's value was undermined as many webmasters sought to garner links purely to influence Google into sending them more traffic, irrespective of whether the link was useful to human site visitors.

It was time for Google—and other search engines—to look at a wider range of off-site factors. There were other reasons to develop more intelligent algorithms. The Internet was reaching a vast population of non-technical users who were often unable to use advanced querying techniques to reach the information they were seeking and the sheer volume and

complexity of the indexed data was vastly different from that of the early days. Search engines had to develop predictive, semantic, linguistic and heuristic algorithms.

A proxy for the PageRank metric is still displayed in the Google Toolbar, but PageRank is only one of more than 100 factors that Google considers in ranking pages.

Today, most search engines keep their methods and ranking algorithms secret. A search engine may use hundreds of factors in ranking the listings on its SERPs; the factors themselves and the weight each carries may change continually.

Much current SEO thinking on what works and what doesn't is largely speculation and informed guesses. Some SEOs have carried out controlled experiments to gauge the effects of different approaches to search optimization.

The following, though, are some of the considerations search engines could be building into their algorithms, and the list of Google patents may give some indication as to what is in the pipeline:

- Age of site
- Length of time domain has been registered
- Age of content
- Regularity with which new content is added
- Age of link and reputation of linking site
- Standard on-site factors
- Negative scoring for on-site factors (for example, a dampening for sites with extensive keyword meta tags indicative of having being SEO-ed)
- Uniqueness of content
- Related terms used in content (the terms the search engine associates as being related to the main content of the page)
- Google Pagerank (Only used in Google's algorithm)
- External links, the anchor text in those external links and in the sites/pages containing those links
- Citations and research sources (indicating the content is of research quality)
- Stem-related terms in the search engine's database (finance/financing)
- Incoming backlinks and anchor text of incoming backlinks
- Negative scoring for some incoming backlinks (perhaps those coming from low value pages, reciprocated backlinks, etc.)
- Rate of acquisition of backlinks: too many too fast could indicate "unnatural" link buying activity
- Text surrounding outward links and incoming backlinks. A link following the words "Sponsored Links" could be ignored
- Use of "rel=nofollow" to suggest that the search engine should ignore the link
- Depth of document in site
- Metrics collected from other sources, such as monitoring how frequently users hit the back button when SERPs send them to a particular page
- Metrics collected from sources like the Google Toolbar, Google AdWords/Adsense programs, etc.

- Metrics collected in data-sharing arrangements with third parties (like providers of statistical programs used to monitor site traffic)
- Rate of removal of incoming links to the site
- Use of sub-domains, use of keywords in sub-domains and volume of content on sub-domains… and negative scoring for such activity
- Semantic connections of hosted documents
- Rate of document addition or change
- IP of hosting service and the number/quality of other sites hosted on that IP
- Other affiliations of linking site with the linked site (do they share an IP? have a common postal address on the "contact us" page?)
- Technical matters like use of 301 to redirect moved pages, showing a 404 server header rather than a 200 server header for pages that don't exist, proper use of robots.txt
- Hosting uptime
- Whether the site serves different content to different categories of users (cloaking)
- Broken outgoing links not rectified promptly
- Unsafe or illegal content
- Quality of HTML coding, presence of coding errors
- Actual click through rates observed by the search engines for listings displayed on their SERPs
- Hand ranking by humans of the most frequently accessed SERPs

The relationship between SEO and the search engines

Search engine operators became interested in the SEO community in the late 1990s. A number of high profile SEO community leaders established contractual relationships with search engines for advertising and consulting purposes. These early contacts led to an amelioration of some hostile feelings between the search optimization and search engineering communities.

In early 2000, search engines and SEO firms attempted to establish an unofficial "truce." There are several tiers of SEO firms, and the more reputable companies employ content-based optimizations which meet with the search engines' (reluctant) approval. These techniques include improvements to site navigation and copywriting, designed to make websites more intelligible to search engine algorithms.

Some search engines have also reached out to the SEO industry, and are frequent sponsors and guests at SEO conferences and seminars. In fact, with the advent of paid inclusion, some search engines now have a vested interest in the health of the optimization community.

Getting into search engines' listings

New sites do not need to be "submitted" to search engines to be listed. A simple link from an established site will get the search engines to visit the new site and spider its contents. It is rarely more than a few days from the acquisition of the link to all the main search engine spiders visiting and indexing the new site.

Once the search engine has found the new site, it will generally visit and index all the pages on the site, as long as all the pages are linked to with standard <a href> hyperlinks. Pages which are accessible only through Flash or Javascript links may not be findable by the spiders.

Webmasters can instruct spiders to not index certain files or directories through the standard robots.txt file in the root directory of the domain. Standard practice requires a search engine to check this file upon visiting the domain. The web developer can use this feature to prevent pages such as shopping carts or other dynamic, user-specific content from appearing in search engine results.

For those search engines who have their own paid submission (like Yahoo), it may save some time to pay a nominal fee for submission.

White hat methods

So-called "white hat" methods of SEO involve following the search engines' guidelines as to what is and what isn't acceptable. Their advice generally is to create content for the user, not the search engines; to make that content easily accessible to their spiders; and to not try to game their system. Often webmasters make critical mistakes when designing or setting up their web sites, inadvertently "poisoning" them so that they will not rank well. White hat SEO attempts to discover and correct mistakes, such as machine-unreadable menus, broken links, temporary redirects, or a generally poor navigation structure that places pages too many clicks from the home page.

Because search engines are text-centric, many of the same methods that are useful for web accessibility are also advantageous for SEO. Methods are available for optimizing graphical content, including ALT attributes, and adding a text caption. Even Flash animations can be optimized by using an OBJECT element that contains equivalent HTML content.

Some methods considered proper by the search engines:

- Using a short and relevant title to name each page.
- Editing web pages to replace vague wording with specific terminology that is relevant to the subject of the page.
- Increasing the amount of original content on a site.
- Using a reasonably-sized, accurate description meta tag without excessive use of keywords, exclamation marks or off topic terms.
- Ensuring that all pages are accessible via regular links, and not only via Java, Javascript or Macromedia Flash applications; this can be done through the use of a page listing all the contents of the site (a Site map)
- Developing links via natural methods: Google doesn't elaborate on this somewhat vague guideline. Dropping an email to a fellow webmaster telling him about a great article you've just posted, and requesting a link, is most likely acceptable.
- Participating in a web ring with other web sites as long as the other websites are independent, share the same topic, and are of comparable quality.

Black hat methods

Spamdexing is the pro motion of irrelevant, chiefly commercial, pages through *deceptive techniques* and the abuse of the search algorithms. Many search engine administrators consider any form of search engine optimization used to improve a website's page rank as spamdexing. However, over time a widespread consensus has developed in the industry as to what are and are not acceptable means of boosting one's search engine placement and resultant traffic.

As search engines operate in a highly automated way it is often possible for webmasters to use methods and tactics not approved by search engines to gain better ranking. These methods often go unnoticed unless an employee from the search engine manually visits the site and notices the activity, or a change in ranking algorithm causes the site to lose the advantage thus gained. Sometimes a company will employ an SEO consultant to evaluate competitor's sites, and report "unethical" optimization methods to the search engines.

Spamdexing often gets confused with legitimate search engine optimization techniques, which do not involve deceit. Spamming involves getting web sites more exposure than they deserve for their keywords, leading to unsatisfactory search results. Optimization involves getting web sites the rank they deserve on the most targeted keywords, leading to satisfactory search experiences.

When discovered, search engines may take action against those found to be using unethical SEO methods. In February 2006, Google removed both BMW Germany and Ricoh Germany for use of these practices.

Legal issues

In 2002, search engine manipulator SearchKing filed suit in an Oklahoma court against the search engine Google. SearchKing's claim was that Google's tactics to prevent spamdexing constituted an unfair business practice. This may be compared to lawsuits which email spammers have filed against spam-fighters, as in various cases against MAPS and other DNSBLs. In January of 2003, the court pronounced a summary judgment in Google's favor.

High quality web sites typically rank well

A webmaster who wants to maximize the value of a web site can read the guidelines published by the search engines, as well as the coding guidelines published by the World Wide Web Consortium. If the guidelines are followed, and the site presents frequently updated, useful, original content, and a few meaningful, useful inbound links are established, it is usually possible to obtain a significant amount of organic search traffic.

When a site has useful content, other webmasters will naturally place links to the site, increasing its PageRank and flow of visitors. When visitors discover a useful web site, they tend to refer other visitors by emailing or instant messaging links.

As a result, SEO practices that improve web site quality are likely to outlive short term practices that simply seek to manipulate search rankings. The top SEOs recommend targeting the same thing that search engines seek to promote: relevant, useful content for their users.

Relevance

In computer science, and particularly in search engines, **relevance** is a numerical score assigned to a search result, representing how well the result meets the information need of the user that issued the search query. In many cases, a result's relevance determines the order in which it is presented to the user.

In academic information retrieval, the word relevance has been used in system evaluation for over forty years, going back to the Cranfield Experiments of the early 1960s. In the relatively new commercial search realm, among web search engine companies, search engine optimizers, and in the press, the incorrect *relevancy* is mistakenly being used more and more instead of the correct *relevance*. One can often tell from which community an information retrieval practitioner hails, depending on whether he or she uses the correct form of the word. Wikipedia's search facility is an example of use of the incorrect *relevancy*.

Algorithms for relevance

In the simplest case, relevance can be calculated by examining how many times a query term appears in a document (term frequency), possibly combined with how discriminative that query term is across the searched collection (often called Term Frequency-Inverse Document Frequency).

Since search engines and other businesses rely upon the accuracy of their results, many additional, more complex algorithms have been developed to estimate result relevance. Many of these algorithms, particularly those used by search engines, are hidden to the public, as a user that knows the details of a search algorithm can artificially boost his own content's ranking.

Relevance calculation is often misinterpreted by the press. For example, it has often been said that when Google burst onto the scene it was miles ahead of its competitors because it, unlike anyone else, ranked web pages by relevance. This is not true since everyone ranks by relevance. It is just that Google had come up with a fairly new way of estimating relevance, namely PageRank. But even search engines that only use TFIDF rank by relevance.

Clustering and relevance

There is also much confusion between the notions of *similarity* and *relevance*. These are not the same thing. It has often been said by many companies doing topic clustering, document filtering, and other such applications that their algorithms function by grouping *relevant* documents together. What is actually meant is that the algorithms are grouping *similar* documents together. Two (or more) documents are never relevant to each other. They may be similar to each other, but they are only ever relevant to a user's information need. If there is no user information need, there is no relevance.

The cluster hypothesis in information retrieval says that two documents that are similar to each other have a high likelihood of being relevant to the same information need. Documents by themselves, however, are never relevant to each other. Relevance is defined in terms of a user's information need.

Keyword density

Keyword density is the percentage of words on a web page that match a specified set of keywords. In the context of search engine optimization keyword density can be used as a factor in determining whether a web page is relevant to a specified keyword or keyword phrase. Due to the ease of managing keyword density, search engines usually implement other measures of relevancy to prevent unscrupulous webmasters from creating search spam through practices such as keyword stuffing.

Keyword stuffing

Keyword stuffing is considered to be an unethical Search engine optimization (SEO) technique. Keyword stuffing occurs when a web page is loaded with keywords in the meta tags or in content. The repetition of words in meta tags, may explain why many search engines no longer use these tags.

Keyword stuffing is used is to obtain maximum search engine ranking and visibility for particular phrases. A word that is repeated too often may raise a red flag to search engines.

Hiding text out of view of the visitor is done in many different ways. Text colored to blend with the background, CSS "Z" positioning to place text "behind" an image – and therefore out of view of the visitor – and CSS absolute positioning to have the text positioned several feet away from the page center, are all common techniques. As of 2005, some of these invisible text techniques can be detected by major search engines.

"Noscript" tags are another way to place hidden content within a page. While they are a valid optimization method for displaying an alternative representation of scripted content, they may be abused, since search engines may index content that is invisible to most visitors.

Inserted text sometimes includes words that are frequently searched (such as "sex") even if those terms bear little connection to the content of a page, in order to attract traffic to advert-driven pages.

Link campaign

Link campaigns are a form of online marketing and is also a method for search engine optimization. A business seeking to increase the number of visitors to its web site can ask its strategic partners, professional organizations, chambers of commerce, suppliers, and customers to add links from their web sites. Typically a link campaign involves mutual links back and forth between related sites.

Increasing the number of links to a site has two beneficial effects:

- Search engines such as Google judge the importance of a site by the number of other sites that link to it.

- The additional links result in visitors moving from the linking site to the target site.

When conducting a link campaign, the essentials steps are to identify potential link partners, request the links, and specify the link text. The value of a link depends on the traffic and reputation of the linking site, and the relevancy of its content to the target site's content. Off topic links are generally not useful because they tend to upset visitors, and search engines may view them as link spam.

Link farms are web sites set up solely for the purpose of exchanging links. These sites are viewed dimly by search engines, and Google specifically advises webmasters not to participate in link farms:

"Linking schemes will often do a site more harm than good."

Link exchange

Link Exchange ("Reciprocal Link Exchange") is the practice of exchanging links with other websites. There are many different ways to arrange a link exchange with webmasters. The simplest way of doing it is to email another website owner and ask to do a link exchange. Also visiting webmaster discussion boards which offer a dedicated link exchange forum where webmasters can request a link exchange be it of a certain category or open to anybody. You place their link on your site, usually on a links page and the other site in return will place a link back to you.

Link exchange has been a long time practice by website owners since the beginning of the WWW. In the last few years (after year 2000), this practice has gained more popularity as search engines such as Google started favoring sites that had more links in the rankings. This system was very accurate at gauging the importance of a website when it first started, leading to the popularity of Google

However according to experts, search engines no longer place a heavy emphasis on reciprocal links. Instead the popularity or credibility of your site is now gauged by one way incoming links to your site. How than do you go about building one way back links to your site? There are a number of proven techniques you can follow:

1. First and foremost your aim should be to link to sites with a similar theme as your site. For example if you site is about "dogs" than it makes sense that back link from another dog or animal related site would be given a heavier weighting that a link from a casino site. You should start by conducting a search with you keywords on the major search engines (MSN, Yahoo, and Google) to come up with a list of sites which appear for that keyword. Next determine the contact info, ideally an e-mail address. Once you have this information, you can simply contact the webmaster (politely) and ask them if they would be willing to link to your site.

2. Another effective way of increasing your link popularity is to write and submit your articles to sites such as articlecity.com. The importance of this is that when you submit your material there is usually a resource box where you can enter the link information to

your site. Every time someone publishes your article, you will have a one way link from their site to yours.

3. Submit to directories under the appropriate category. Many directories and human edited and therefore a link from a directory can instantly add credibility to your site. A major directory is Dmoz. Since site submissions are human reviewed, expect at least 6-8 weeks for any kind of response.

4. Submit your URL to link exchange directories where web users such as your self are actively looking to find new relevant link partners. If you search for google.com, ask.com or msn.com for terms: link exchange or link trade you will be able to find some good ones.

Reciprocal link

A **reciprocal link** is a mutual link between two objects, commonly between two websites in order to ensure mutual traffic. Example: Alice and Bob have websites. If Bob's website links to Alice's website, and Alice's website links to Bos's website, the websites are reciprocally linked.

Website owners often submit their sites to reciprocal link exchange directories in order to achieve higher rankings in the search engines.

Link farm

On the World Wide Web, a **link farm** is any group of web pages that all hyperlink to every other page in the group. Although some link farms can be created by hand, most are created through automated programs and services. A link farm is a form of spamming the index of a search engine (sometimes called spamdexing).

History

Link farms were developed by search engine optimizers in 1999 to take advantage of the Inktomi search engine's dependence upon link popularity. Although link popularity is used by some search engines to help establish a ranking order for search results, the Inktomi engine at the time maintained two indexes. Search results were produced from the primary index which was limited to approximately 100,000,000 listings. Pages with few inbound links continually fell out of the Inktomi index on a monthly basis.

Inktomi was targeted for manipulation through link farms because it was then used by several independent but popular search engines, such as HotBot. Yahoo!, then the most popular search service, also used Inktomi results to supplement its directory search feature. The link farms helped stabilize listings for (normally) online business Web sites that had few natural links from larger more stable sites in the Inktomi index.

Link farm exchanges were at first handled on an informal basis, but several service companies were founded to provide automated registration, categorization, and link page updates to member Web sites.

When the Google search engine became popular, search engine optimizers learned that Google's ranking algorithm depended in part on a link weighting scheme called PageRank. Rather than simply count all inbound links equally, the PageRank algorithm determines that some links may be more valuable than others, and therefore assigns them more weight than others. Link farming was adapted to help increase the PageRank of member pages.

However, even the link farms became susceptible to manipulation by unscrupulous Webmasters who joined the services, received inbound linkage, and then found ways to hide their outbound links or to avoid posting any links on their sites at all. Link farm managers had to implement quality controls and monitor member compliance with their rules to ensure fairness.

Alternative link farm products emerged, particularly link-finding software that identified potential reciprocal link partners, sent them template-based emails offering to exchange links, and create directory-like link pages for Web sites hoping to build their link popularity and PageRank.

Search engines countered the link farm movement by identifying specific attributes associated with link farm pages and filtering those pages from indexing and search results. In some cases, entire domains were removed from the search engine indexes in order to prevent them from influencing search results.

Justification

The justification for link farm-influenced crawling diminished proportionately as the search engines expanded their capacities to index more sites. Once the 500,000,000 listing threshold was crossed, link farms became unnecessary for helping sites stay in primary indexes. Inktomi's technology, now a part of Yahoo!, now indexes billions of Web pages and uses them to offer its search results.

Where link weighting is still believed by some Webmasters to influence search engine results with Google, Yahoo!, MSN, and Ask (among others), link farms remain a popular tool for increasing PageRank or perceived equivalent values. PageRank-like measurements apply only to the individual pages being linked to (typically the reciprocal linking pages on member sites), so these pages must in turn link out to other pages (such as the main index pages of the member sites) in order for the link weighting to help.

The expression "link farm" is now considered to be pejorative and derogatory. Many reciprocal link management service operators tout the value of their resource management and direct networking relationship building. The reciprocal link management services promote their industry as an alternative to search engines for finding and attracting visitors to Web sites. Their acceptance is by no means universal but the link management services seem to have established a stable customer base.

Guidelines

Google indicates in its Webmaster Guidelines that more than 100 factors are used to determine search results rankings. There is considerable debate in the search engine optimization community regarding the continued value of PageRank. Mike Grehan, a well-

known search engine optimization columnist, has publicly quoted engineers from Yahoo! and Ask who say Google never fully implemented their PageRank algorithm.

Search engines such as Google recommend that webmasters request **relevant** links to their sites (conduct a link campaign), but avoid participating in link farms. According to Google, a site that participates in a link farm may have its search rankings penalized.

Search engines try to identify specific attributes associated with link farm pages and filter those pages from indexing and search results. In some cases, entire domains are removed from the search engine indexes in order to prevent them from influencing search results.

Link popularity

Link popularity is a measure of the quantity and quality of other web sites that link to a specific site on the World Wide Web. It is an example of the move by search engines towards off-the-page-criteria to determine quality content. In theory, off-the-page-criteria adds the aspect of impartiality to search engine rankings.

Link popularity plays an important role in the visibility of a web site among the top of the search results. Indeed, some search engines require at least one or more links coming to a web site, otherwise they will drop it from their index.

Search engines such as Google use a special link analysis system to rank web pages. Citations from other WWW authors help to define a site's reputation. The philosophy of link popularity is that important sites will attract many links. Content-poor sites will have difficulty attracting any links. Link popularity assumes that not all incoming links are equal, as an inbound link from a major directory carries more weight than an inbound link from an obscure personal home page. In other words, the quality of incoming links counts more than sheer numbers of them.

To search for pages linking to a specific page, simply enter the URL on Google or Yahoo! this way:

link: http://yourdomainname/pagename.html

Here are some strategies that are generally considered to be important to increase link popularity:

- There should be links from the home page to all subpages so that a search engine can transfer some link popularity to the subpages.
- Appropriate anchor text with relevant keywords should be used in the text links that are pointing to pages within a site (technically, this helps link context, not link popularity).
- Getting links from other web sites, particularly sites with high PageRank, can be one of the most powerful site promotion tools. Therefore, the webmaster should try to get links from other important sites offering information or products compatible or synergistic to his/her own site or from sites that cater to the same audience the webmaster does. The webmaster should explain the advantages to the potential link partner and the advantages his/her site has to their visitors.

- One way links often count for more than reciprocal links.
- The webmaster should list his/her site in one or more of the major directories such as Yahoo! or the Open Directory Project.
- The webmaster should only link to sites that he/she can trust, i.e. sites that do not use "spammy techniques".
- The webmaster should not participate in link exchange programs or link farms, as search engines will ban sites that participate in such programs.

To increase link popularity, many webmasters interlink multiple domains that they own, but often search engines will filter out these links, as such links are not independent votes for a page and are only created to trick the search engines. In this context, closed circles are often used, but these should be avoided, as they hoard PageRank.

Anchor text

Anchor text is the visible text in a hyperlink. Anchor text is weighted (ranked) highly in search engine algorithms, because the linked text is usually relevant to the landing page. The objective of search engines is to provide highly relevant search results; this is where anchor text helps, as the tendency is, more often than not, to hyperlink words relevant to the landing page.

Usually this is exploited by webmasters to procure high results in SERPS (search engine results pages). Google Bombing is possible through anchor text manipulation. Much has been written on anchor text which is available on the web today.

Although the search engines are well aware of anchor text manipulation, not much change can be expected in the SE algorithms in the near future because the brighter side of the picture cannot be overlooked: anchor text delivers relevance.

The latest changes in Google's algorithm point towards discounting of websites in the SERPS which are involved in anchor text manipulation for higher rankings. Although there are no indications of this clause (of the latest Google patent) having been implemented, this could become a reality in the future.

Site map

A **site map** (or **sitemap**) is a web page that lists the pages on a web site, typically organized in hierarchical fashion. This helps visitors, and search engine robots, to find pages on the site. An example is the one on EFF's (Electronic Frontier Foundation's) page.

Site maps can improve search engine optimization of a site by making sure that all the pages can be found. This is especially important if a site uses Macromedia Flash or JavaScript menus that do not include HTML links.

Site maps do have limitations. Most search engines will only follow a finite number of links from a page, so if a site is very large, additional strategies besides the site map may be required that search engines, and visitors, can access all content on the site.

While some developers argue that **site index** is a more appropriately used term to relay page function, web visitors are used to seeing each term and generally associate both as one in the same.

Google Sitemaps

Google maintains a feature called Google Sitemaps that allows web developers to publish lists of links from across their sites. The basic premise is that some sites have a large number of pages that are only available through the use of forms and user entries. The sitemap files can then be used to indicate to a web crawler how such pages can be found.

Search engine results page

A **search engine results page**, or **SERP** is the listing of web pages returned by a search engine in response to a keyword query. The results normal include a list of web pages with titles, a link to the page, and a short description showing where the keywords have matched content within the page. SERPs may also contain advertisements, a way for search engines to earn income.

For frequently requested search terms, search engines do not build the SERP afresh each time. Instead, the search engines build the SERP once and save it for subsequent queries. Periodically the search engine crawls the web and rebuilds the SERP to account for new pages, and to re-rank the pages according to the search engine's own algorithm.

As of 2006, SERPs are offering search engine users more features, including local matches, files on the user's own computer, definitions, images, and suggestions for alternative keywords.

Search engine optimization is a sub-specialty within web development that seeks to position a web site more frequently and more prominently in the SERPs. Spamdexing is the use of deceptive practices to raise a web site's visibility in the SERPs.

Organic search

An **organic search** is a process by which World Wide Web users find web sites having unpaid search engine listings, as opposed to using the pay per click advertisement listings displayed among the search results.

The field of search engine optimization, (SEO), is concerned with maximizing the visibility of a web site by making its listings appear more frequently and more prominently in organic search results. Some businesses base their SEO strategies on the successful insertion of their web site listing(s) into organic search results for various popular keywords.

P4P

P4P is an abbreviation of the term "Pay For Performance". The concept was invented at Overture (now Yahoo! Search Marketing) and later adopted by their competitors, most famously Google's AdWords. Under the model advertisers bid on the rights to present a search result for a specific search terms in an open auction. When someone enters a search term that has been bid on, the results from the auction on that search term are presented, ranked from highest bid to lowest.

When this idea was first introduced, there was a lot of controversy. Many felt that the results would be irrelevant, but the auction model has proven to be quite effective at producing relevant results for searches where you want to buy something. P4P is now seen as the most efficient and effective way to monetize search engines.

Paid inclusion

Paid inclusion is a search engine marketing product where the search engine company charges fees related to inclusion of websites in their search index. Paid inclusion products are provided by most search engine companies, the most notable exception being Google.

The fee structure is both a filter against superfluous submissions and a revenue generator. Typically, the fee covers an annual subscription for one webpage, which will automatically be catalogued on a regular basis. A per-click fee may also apply. Each search engine is different. Some sites allow only paid inclusion, although these have had little success. More frequently, many search engines, like Yahoo!, mix paid inclusion (per-page and per-click fee) with results from web crawling. Others, like Google, do not let webmasters pay to be in their search engine listing (advertisements are shown separately and labeled as such).

Some detractors of paid inclusion allege that it causes searches to return results based more on the economic standing of the interests of a web site, and less on the relevancy of that site to end-users.

Often the line between pay per click advertising and paid inclusion is debatable. Some have lobbied for any paid listings to be labeled as an advertisement, while defenders insist they are not actually ads since the webmasters do not control the content of the listing, its ranking, or even whether it is shown to any users.

Paid inclusion is a search engine marketing method in itself, but also a tool of search engine optimization, since experts and firms can test out different approaches to improving ranking, and see the results often within a couple of days, instead of waiting weeks or months. Knowledge gained this way can be used to optimize other web pages, without paying the search engine company.

Google consultant

A **Google consultant** is a person or company specializing in search engine optimization (SEO) for the Google search.

Because Google is the most widely used internet search engine the commercial importance of achieving a good page rank on Google has grown hugely for many companies - particularly those businesses that rely on internet marketing to attract visitors or customers to their websites.

While it is possible for companies to buy position on Google using the Google AdWords system the results tend not to be regarded by those conducting the search as particularly authentic or as unbiased as the page rankings. Also, it can be very expensive to advertise with popular keywords.

As Google's popularity increased companies tried improving their rankings first by using Google bombing in which they created many different sites all linked to one another through a particular word. This gave the site a high ranking when the word was searched for. This "abuse" threatened the utility of Google as a search engine so Google responded by adapting its top secret web crawler technologies and ranking algorithms.

In turn SEO consultants tried to reverse engineer the Google technology in order to gain advantage in the page rankings. So started a struggle between Google and SEO consultants.

As it became more and more difficult to trick the Google system—so the Google consultants became more and more specialised in their work to improve the Google rankings of their clients (eventually to the point where they don't work on any other search engine). The careful use of copywriting, linking strategies and the frequency and timing of updated content can all help move sites up the rankings; the closer a site is to the top of the first page of results delivered by Google the more likely it is that the person searching will click on the link.

The boom in popularity of personal Blogs and Blogging has also had an impact on the way sites are ranked; this is reflected in the purchase of blogger.com by Google

Google and the Google Consultants have more or less reached a détente where Google gives a certain amount of advice to Google consultants on how best to edit or engineer webpages to allow them to be indexed properly. However since Google is now listed on NASDAQ and its commercial success depends to a great extent on selling AdWords by auction it remains to be seen if the symbiotic relationship between Google and the Google consultants will survive.

Google bomb

A **Google bomb** or **Google washer** is a certain attempt to influence the ranking of a given page in results returned by the Google search engine, often with humorous intentions. Due to the way that Google's PageRank algorithm works, a page will be ranked higher if the sites that link to that page all use consistent anchor text. A Google bomb is created if a large

number of sites link to the page in this manner. *Google bomb* is used both as a verb and a noun.

See Spamdexing for the practice of deliberately and dishonestly modifying HTML pages to increase the chance of them being placed close to the beginning of search engine results, or to influence the category to which the page is assigned in a dishonest manner.

Background

An example of Google bombing is if a user registers many domains and all of them link to a main site with the text "... is a living legend". Searching for "living legend" on Google will return the main site higher in the ranking, even if the phrase "living legend" doesn't appear on the main site. A common means of exploiting this is through weblogs, where although the entry may disappear from the main page quickly, the short-term effects of a link can dramatically affect the ranking of a given site. Empirical results indicate that it does not take a large number of websites to achieve a Googlebomb. The effect has been achieved with only a handful of dedicated weblogs.

The above has to be qualified, however. A handful of blog links will not Google bomb someone like Amazon.com out of the top results for "books," for example. In fact, Googlebombs have generally had an impact on relatively "non-competitive" terms, where there's no particular page that seems to be necessarily the right answer.

The technique was first discussed on April 6, 2001 in an article by Adam Mathes. In that article, he coined the term "Google bombing" and explained how he discovered that Google used the technique to calculate page rankings. He found that a search for "internet rockstar" returned the website of a Ben Brown as the first result, even though "internet rockstar" did not appear anywhere on Brown's webpage. He reasoned that Google's algorithm returned it as the first result because many fan sites that linked to Brown's website used that phrase on their own pages.

Mathes began testing his theory by setting out to make the website of his friend Andy Pressman the number one result for a query of "talentless hack". He gave instructions for creating websites and links to Pressman's website with the text of the link reading "talentless hack". Sure enough, as other webloggers joined in his Googlebombing campaign, Pressman's website became the number one result in a Google search for "talentless hack." (By 2004, Mathes's own site was the number one Google result of this search term.)

However, the first Google bomb mentioned in the popular press may have occurred accidentally in 1999, when users discovered that the query "more evil than Satan" returned Microsoft's home page. Now, it returns links to several news articles on the discovery.

Google bombs often end their life by being too popular or well known, thereby attaining a mention in well-regarded web journals and knocking the bomb off the top spot. It is sometimes commented that Google bombing need not be countered because of this self-disassembly.

In addition, the entire notion of "Google bombs" might be better described as "link bombing," given that these campaigns can certainly have an effect on other search engines, as well. All major search engines make use of link analysis and thus can be impacted. So, a

search for "miserable failure" on June 1, 2005 brought up the official George W. Bush biography number one on Google, Yahoo and MSN and number two on Ask Jeeves. On June 2, 2005, Yooter reported that George Bush is now ranked first for the keyword 'failure' as well as 'miserable failure' in both Google and Yahoo. And on September 16, 2005, Marissa Mayer wrote on Google Blog about the practice of Google bombing and the word "failure." (See Google's response below) Other large political figures have been targeted for Google bombs such as, Yooter reported on January 6, 2006, Tony Blair is now indexed in the US version of Google for the keyword 'liar'.

The BBC in reporting on Google bombs in 2002 actually used the headline of "Google Hit By Link Bombers," acknowledging to some degree the idea of "link bombing." In 2004, the Search Engine Watch site said that the term should be "link bombing" because of the impact beyond Google and continues to use that term as more accurate.

Nevertheless, "Google bombing" was added to the New Oxford American Dictionary in May 2005.

Googlebombing competitions

In May 2004, Dark Blue and SearchGuild.com teamed up to create what they termed the "SEO Challenge".

The contest sparked controversy around the Internet, as some groups worried that search engine optimization (SEO) companies would abuse the techniques used in the competition to alter queries more relevant to the average user. This fear was offset by the belief that Google would alter their algorithm based on the methods used by the googlebombers.

In September 2004, another SEO contest was created. This time, the objective was to get the top result for the phrase "seraphim proudleduck". A large sum of money was offered to the winner, but the competition turned out to be a hoax.

In .net magazine, Issue 134, March 2005, a contest was created among five professional web site developers to make their site the number one listed site for the made-up phrase "crystalline incandescence".

Google's response

Google has defended its algorithms as simply a reflection of the opinion on the Internet, saying that they are not damaging the overall quality of its services. Google has said it expects Googlebombing to return to obscurity and has dismissed it as "cybergraffiti" and just another internet fad.

On 18 January 2005 the Google blog entry "Preventing comment spam" declared that Google will henceforth respect a `rel="nofollow"` attribute on hyperlinks. Their page ranking algorithm now avoids links with this attribute when ranking the destination page. The intended result is that site administrators can easily modify user-posted links such that the attribute is present, and thus an attempt to googlebomb by posting a link on such a site would yield no increase in that link's rank.

On 16 September 2005 Marissa Mayer, Director of Consumer Web Products for Google wrote an entry on Google Blog to those who were offended by the result of President

George W. Bush's biography with the search of "failure", "miserable", and "miserable failure", stating that Google has no control over and does not condone the act of Google bombing. Apparently, people who sent in complaints believed that the search results showed Google's political bias.

We don't condone the practice of googlebombing, or any other action that seeks to affect the integrity of our search results, but we're also reluctant to alter our results by hand in order to prevent such items from showing up. Pranks like this may be distracting to some, but they don't affect the overall quality of our search service, whose objectivity, as always, remains the core of our mission.

Googlebombing in general

In some cases, the phenomenon has produced competing attempts to use the same search term as a Googlebomb. As a result, the first result at any given time varies, but the targeted sites will occupy all the top slots using a normal search instead of "I'm feeling lucky". Notable instances of this include failure and miserable failure. The primary targets have been the Bush biography above, and Michael Moore's website at www.michaelmoore.com.

Searching for miserabile fallimento (Italian for "miserable failure") was at one time returning Berlusconi biography, but as of 4th January 2006 returns a report on the phenomenon from an Italian news website, *portalino.it*. The words *raar kapsel* (Dutch for "funny hair style") return Jan Peter Balkenende's (Dutch prime minister) biography.

Other search engines use similar ways to rank results, so Yahoo!, AltaVista, and HotBot are also affected by Google Bombs. A search of "miserable failure" on the aforementioned search engines produces the biography of George W. Bush listed at the White House site as the first link on the list. Only a few search engines, such as Ask Jeeves!, MetaCrawler and ProFusion, do not produce the same first links as the rest of the search engines. MetaCrawler and ProFusion are metasearch engines which use multiple search engines.

Googlebombing as Political Activism

Obviously, some of the most famous google bombs are also expressions of political opinion (e.g. "liar" leading to Blair or "failure" leading to Bush.) In general, one of the keys to Google's success has been its ability to capture what ordinary web citizens believe to be important via the information provided in webpage links. One of Google's failures has been their inability to stop organized / commercial exploitation of their algorithms. But is a googlebomb an exploitation, or a democratic expression of opinion? Does organization in itself imply misuse?

One extremely successful, long-lasting and widespread link bomb has been the linking of the term "Scientology" to Operation Clambake. In this case, the index rating clearly emerges from both the individual decisions of pagewriters and reporters and an organized effort lead by the Operation Clambake itself. In this case, the "bombers" believe they may be saving people's lives by giving them important information. The Church of Scientology has also sometimes been accused of an attempt at googlebombing for making a large number of websites linking terms "Scientology" and "L. Ron Hubbard" to each other.

A google bomb could be achieved easily, this is a possible scenario:

- The initiator chooses a word to be searched : "liars"
- The initiator chooses the target website : "http://example.com/"
- The initiator creates a link like this : `<a href="http://example.com/">liars</a>`
- The initiator places this code in his website, as his signature in forum, in his blogs etc.
- The initiator talks to other people about the bomb and tells other people to use the code in their own writings.
- GoogleBot indexes and ranks, resulting in http://www.google.co.uk/search?num=100&hl=en&newwindow=1&q=liars&meta= having the political party's webpage as a first result

There is a blog about google bombs, that people wanting to initate a google bomb visit, and submit their bombs. People watching this blog include the bombs in their blogs. : http://gbombs.blogspot.com/ This blog has been used in the past to initate political bombs mainly, in many countries.

Commercial googlebombing

Some unscrupulous website operators have adapted googlebombing techniques to do spamdexing. This includes, among other techniques, posting of links to a site in an Internet forum along with phrases the promoter hopes to associate with the site (see Spam in blogs). Unlike conventional message board spam, the object is not to attract readers to the site directly, but to increase the site's ranking under those search terms. Promoters using this technique frequently target forums with low reader traffic, in hopes that it will fly under the moderators' radar. Wikis in particular are often the target of this kind of page rank vandalism, as all of the pages are freely editable.

Another technique is for the owner of an Internet domain name to set up the domain's DNS entry so that all subdomains are directed to the same server. The operator then sets up the server so that page requests generate a page full of desired Google search terms, each linking to a subdomain of the same site, with the same title as the subdomain in the requested URL. Frequently the subdomain matches the linked phrase, with spaces replaced by underscores or hyphens. Since Google treats subdomains as distinct sites, the effect of a large number of subdomains linking to each other is a boost to the PageRank of those subdomains and of any other site they link to.

As of 2 February 2005, many have noticed changes in the Google algorithm that largely affects, among other things, Googlebombs. As evidence of this, ponder that only roughly 10% of the googlebombs listed below work as of 15 February 2005. This is largely due to Google refactoring its valuation of PageRank, mostly in an effort to keep up with the encroaching result relevancy of the Yahoo and MSN search engines, which many people claim are not nearly as easy to "hack" as Google.

The Quixtar Google bombing example

In 2005, multi-level marketing giant Quixtar began a "Quixtar Web Initiative" to manipulate Google results. The project was considered by many to be a clear and flagrant violation of Google's Quality Guidelines.

According to a web article called Quixtar Admits Google Bombing, a Quixtar Diamond told members that the company had "hired geekoids who were spending their time Google bombing positive info about Quixtar so that the negative sites would be buried way down at the bottom of the Google list when a prospect types in Quixtar [in a search engine]. Nobody will even be able to find the negative sites anymore."

The goal presented was to smother anti- Quixtar sites, such as a consumer protection group, an eBook about Amway and Quixtar, and grassroots movements from ex-Quixtar members that claim Amway and Quixtar are Pyramid schemes, cults, and use deceptive business practices.

Quixtar's attempt to lower the ranking of such negative web sites backfired, and, for a few months, Quixtar's Google results suffered for its attempted Googlebombing.

As of February 2006, a search on Google for "Quixtar " reveals that Quixtar.com is again the number one result.

The organized effort to manipulate search engine results was first brought to light by bloggers which detailed the evidence in an article "So Busted".

Ironically, the site that reported Quixtar's attempt to Google bomb once occupied the first search position under "Quixtar". Often, in fact, a majority of the top ten results for the term "Quixtar" are sites critical of the Quixtar business opportunity and its lines of sponsorship. This is mixed in with real positive results, however, and helps to show both sides of the arguments.

Quixtar's initiative included at least 54 Quixtar Blogs and numerous other company-related sites, character assassination blogs, adoration blogs, and various other pages. Immediately after the exposure many of the blogs shut down or reduced their content.

Some of Quixtar's independent lines of sponsorship have also been among the largest abusers of Google bombing. These independent lines of sponsorship are led by high-level Independent Business Owners, or IBOs, called Diamonds, Crowns, and Crown Ambassadors.

Search engine bombing before Google

Before Google existed, eccentric USENET poster Archimedes Plutonium, upset with the attention he received from users who found him amusing, posted an angry message to two science newsgroups. He accused these people of "SearchEnginebombing," an offshoot of Emailbombing, that was cluttering the web/USENET with negative comments about him, so a search engine would find more of them than his own postings. Unlike "Google Bombing", the term "Search Engine Bombing" didn't immediately catch on, and initially its use has been primarily limited to Archimedes Plutonium, and USENET posters who mocked him.

Accomplished Googlebombs

Note that Google results fluctuate regularly and the bombed results may move up and down. Most often they will be displaced by a news article describing the bomb.

Recent (as of 2006) and popular examples are:

- "Arabian Gulf" - points to an error look-alike page saying that "the gulf you are looking for does not exist." The page links to the Wikipedia entry on the Persian Gulf, alternative English name for the body of water east of the Arabian Peninsula and south of Iran.
- Aweonao (chilean slang for asshole) in the Google Image Search service shows chilean right-wing presidential candidate, Joaquin Lavín.
- Awful Announcer brings you to the official website of frequently criticized baseball color commentator Tim McCarver
- Bill Napoli brings up an extremely negative definition of South Dakota Republican Senator Bill Napoli. It's based on comments he made after a bill in the South Dakota State Legislature was passed that, if it stands up to the scrutiny of the United States Supreme Court, will ban abortion in that state.
- "bunch of luddites" - points to the homepage of the Motion Picture Association of America.
- Buffone - unofficial Silvio Berlusconi (Italian Prime Minister) biography. "buffone" is the Italian word for "clown". As of 30th January 2006 this is now 2nd in the list.
- Despota Cachaceiro ("Drunk Ruler" in portuguese) used to return the website of Luis Inácio Lula da Silva, brazilian president. This googlebomb emerged by the same time New York Times published a polemic article regarding Lula's allegedly heavy-drinking habits .
- Delincuente ("Criminal" in Spanish) - Compromiso para el Cambio's webpage appears.
- Estupidez - Points to former Ecuadorian President Lucio Gutierrez's webpage. The page is no longer supported, though the link still exists.
- Estúpido Populista - Points to mexican presidential candidate Andres Manuel Lopez Obrador Home page.
- Failure - see "miserable failure" below.
- First against the wall - Used to return the Wikipedia entry on Karl Rove.
- Food Nazis - Points to "Center for Science in Public Interest" (which advocates strict regulatory oversight of genetically engineered foods).
- French military victories - see French military victories (practical joke). Points to a faked Google page implying there are no French military victories.
- fuckwit used to return John Prescott who is Deputy Prime Minister of the United Kingdom, well known for his speech impairment.
- gastrointestinal dysentery returns Kres Chophouse & Lounge in Orlando, Florida, a restaurant that fired a server for blogging about work.
- Gladjakker ("smoothie") returns website of Camiel Eurlings, the leader of the Dutch Christian-Democrat fraction in the European Parliament.
- Great President points to the biography of George W. Bush at the official White House website.
- hell used to put Microsoft's homepage in the top spot. Searching for hell google microsoft returns reference to this.
- Ignorant Asshole returns the website of Cal Thomas. The reason was his column.
- ignorant bigots - returns the official page of Christian Voice, a fundamentalist Christian organisation in the UK.

- International Sign for Choking in Google Images search use to yield the logo of the Philadelphia Eagles football team, shortly after their defeat in Super Bowl XXXIX. The same search recently showed a logo of the New York Yankees baseball team, in reference to their stunning ALCS loss to the Red Sox in 2004. (As of June 2005, neither of these results are returned)
- jämmerlicher Waschlappen (German for "miserable washcloth") returns the government page of Christoph Blocher, a Swiss Federal Councillor.
- Jew - JewWatch, an anti-semitic website operated by Frank Weltner, a white power nationalist, was the number one hit when searching on Google.com for "Jew" in 2005. In early 2006, the Wikipedia entry replaced it following a Googlebombing campaign organized by Daniel Sieradski, editor of the blog Jew School. [3] [4] [5] As of 02 March 2006 the Wikipedia site is again in the number one position, however it appears to be fluctuating regularly. Google added an explanation page entitled Offensive Search Results and placed it in the top sponsored link section.
- kretyn (which means "cretin" in Polish) - returns the page with information about Polish politician Andrzej Lepper
- Iznogoud (a character who wants to become number one) - Points to "Biographie - Nicolas SARKOZY", and "Nicolas Sarkozy" points to "Iznogoud, the movie".
- Ladrones (Spanish for "thieves") and Siempre Ganamos Algunos Euros ("We Always Earn Some Euros") point to the homepage of SGAE, (Sociedad General de Autores y Editores), the Spanish equivalent of the Recording Industry Association of America (RIAA). The SGAE is an extremely unpopular association in Spain since they not only try to prosecute users of P2P applications, but also impose surcharges on the price of physical media such as recordable CDs in order to account for the theoretical losses due to P2P exchanges.
- lažnivec (Slovenian for "liar") points to the page of Bojan Požar, Slovenian yellow press publicist.
- liar on google.co.uk - returned Tony Blair, the UK Prime Minister, accused of misleading the public over weapons of mass destruction in Iraq. (as of 2005-12-30 Blair is back in first place, with the IMDb page for "Liar Liar" falling back into second place) He is also first in the world rankings at google.com on top.
- litigious bastards used to give the homepage of the SCO Group, which initiated the SCO v. IBM lawsuit alleging copyright violation in the Linux kernel. bastards also worked.
- lying sack of shit Used to return the Parliamentary web page of Australian Federal Attorney General Phillip Ruddock. This bomb was suggested by weezil and executed by a number of Australian bloggers in protest of Ruddock's criticism of Mamdouh Habib
- lul (Dutch for 'dick') used to return the web page of the Belgian politician Hugo Coveliers on google.be, after which Coveliers went to the Federal Computer Crime Unit in Belgium.
- Miserabile fallimento - official Silvio Berlusconi (Italian Prime Minister) biography. "miserabile fallimento" is the Italian for "miserable failure".
- Miserable failure, miserable worst president worst president ever and great president brings up the official George W. Bush biography from the US White House

web site. Due to the search query of "miserable failure," the search terms miserable and failure (each word that comprises miserable failure used on their own) also point to the biography of George W. Bush, with Michael Moore ranking number two. Unelectable points to the biography on the White House's homepage. Interestingly enough, www.unelectable.com used to point to the same page, and was second in Google's Search for "unelectable." See also miserable failure. With the addition of Google Local and Maps, searching for the phrase in Washington DC provides George W. Bush's residence (Listed as the "US Executive Mansion") as the first result.

- Mouton insignifiant (French for "trivial sheep") - returns the official page of Jean Charest, Premier of the province of Quebec, in Canada. It refers to his curly hair. Insignifiant also worked
- National Disgrace - returns the official MLB biography of Bud Selig, commissioner of Major League Baseball. The Googlebomb was organized to highlight MLB's poor behavior in the process of moving the Montreal Expos to Washington, D.C. and their hard-line stadium lease negotiation tactics.
- Old Rice And Monkey Nuts returns the website of Herald Sun columnist Andrew Bolt. The phrase is an obscure reference to Tirath Khemlani, a Pakistani commodities trader who was involved in brokering an improbable US$4 billion loan deal to the Australian Government under Prime Minister Gough Whitlam in 1974. Khemlani was known derisively by his usual line of trade - rice and monkey nuts. As he was involved in commodities and not financial transactions as a rule, it was believed that Khemlani did not have access to the funds as he claimed but would attempt to oblige the Australian Government of the day to pay a huge commission for arranging the proposed loan. The bomb was perpetrated at the suggestion of Ausculture for reasons unknown.
- Opportunist - Links to the web site for the former leader of the opposition in the UK, Michael Howard.
- "Out of Touch Executives" - Used to lead to Google's own corporate information page. "Out of Touch Management" used to work as well.
- Pekeng Pangulo or "fake president" in Filipino returns the official page of Philippine President Gloria Macapagal Arroyo.
- Searching UK domains only for poodle gives you a link to a Tony Blair biography (dropped to sixth at one stage but has now returned to the #1 spot).
- populista (Slovak for "populist") - returns an official homepage of Robert Fico, a left-wing Slovak politician.
- purge princess brought up the Senate Campaign blog of Katherine Harris.
- ¿Quién quiere estafarnos? (Spanish for "Who wants to swindle us?") points to the homepage of Telecom, provider of phone and ADSL services in southern Argentina. This bombing was started at http://bombardeo.blogspot.com because of the company's announcement to limit download transfer to 4 GB per month (for 512 kbit/s connections).
- Raar Kapsel ("Weird Haircut" in Dutch) - Returns the biography of the Prime Minister of the Netherlands, Jan Peter Balkenende, who is known for his distinctive hairstyle.

- Santorum - "Spreading Santorum," a campaign to ridicule Senator Santorum by naming a mixture of bodily substances after him.
- Scottish Raj returns the website of Gordon Brown, UK Chancellor of the Exchequer. This googlebomb was initiated by The Campaign for an English Parliament news blog, whose author objected to Brown's calls for a renewed sense of Britishness and ambition to become UK Prime Minister when his native Scotland has its own parliament.
- Siedziba szatana (Polish for "Satan's seat") returns the website of Radio Maryja, Polish ultra-Catholic religious and political radio station.
- swivel eyed loons returns the homepage of the UK Independence Party after the phrase - initially used to describe the party by blogger Anthony Wells - was adopted by several British bloggers.
- Talentless hack typed into Yahoo's search engine once pointed to singer Ashlee Simpson. As of April 2006, Ashlee Simpson ranks eleventh in searches for "talentless hack." The official website of the band Creed (band) is third. The tenth is a link to Jerry Bruckheimer Films.
- terrorist sympathizer returns the homepage of Bill O'Reilly, in reference to his comment that "every other place in America is off limits to [terrorists], except San Francisco." Initiated by Daily Kos.
- Totalt fiasko - Used to return the official Göran Persson (Swedish Prime Minister) biography. The phrase is Swedish for "miserable failure".
- Tyhmä lehmä, Finnish for "stupid cow", used to brought up home page of Tanja Karpela, Finnish Culture Minister.
- Velky bratr (Czech for "Big Brother") - returned a biography of Stanislav Gross, former Prime Minister of the Czech Republic.
- völlige Inkompetenz, German for "total incompetence", returns the homepage of Karl-Heinz Grasser, the Austrian minister of finance.
- Vreemdelingenhaat ("Hate for foreign people" in Dutch) - Returns the biography of the Minister of Integration and Immigration, Rita Verdonk, whose policies are controversial.
- Waffles - Used to lead to John Kerry's 2004 election site, originated here. A play on Kerry's opponents' accusations that he routinely changed back and forth (or "waffled") between various political positions as convenient. However, the first result on the second page (Jan 31, 2006) still leads to John Kerry's official site.
- Weapons of mass destruction - Internet Explorer Error look-alike joke page saying "weapons of mass destruction cannot be found". (Note: as of 25 March 2006, the joke page, although still available at [6], had fallen to 132nd place in the Google search result.)
- Similarly, Armas de destruccion masiva, Spanish for "Weapons of mass destruction", gives a Spanish version of the page above.
- Ληστές, (Greek for "thieves") links to OTE, the Greek Telecommunications Organization. This is the result of a mass blogger protest against OTE's abusive charges.
- ψεύτες, (Greek for "liars") linked to New Democracy, a center-right wing Greek polical party that is in the government of Greece at present.

- ατσαλάκωτος, (Greek for "non creased") linked to the mayor's office of Municipality of Thessaloniki, a profile page for center-right wing New Democracy mayor Vasillis Papageorgopoulos of Thessaloniki, Greece at present.
- Miserable, (Spanish for "miserable") links to the Official Biography of ex-minister Angel Acebes.
- Masendav - If you search for the word "masendav" (frustrating in estonian) you will get to the estonian center pary page. BTW: this used to work allso, when you typed "masendav" in firefox addres bar before [7] came.
- Todo Goiano é corno - a google bomb made by Brazilians based on the quoted regional joke to indulge Google suggests "Todo Baiano é corno".
- Найти ближайший туалет, (Russian for "find the nearest restroom") - used to have Russian McDonald's addresses as the first link, but now the article about the google bomb is first. It is common in Russia to link to McDonald's as a place where one can always find a restroom when no others are around or require an entrance fee. The link to McDonald's however still appears on the first search results page.
- March 20th, will bring you to the new 'Official' men's holiday of Mar 14th...Steakandblowjob Day.

Justice bomb

A **Justice Bomb** (used as both a noun and a verb) is a form of Google Bomb that is created to counter a commercial or perceivedly offensive result on Google.

One of the most high profile Justice Bombs was the "Joogle Bomb" which in 2004 was carried out in order to oust Jew Watch, an anti-semitic site, from its status as the #1 Google result for the word "Jew".

Google juice

Google juice is jargon for the ability or power of a website to turn up in Google searches. A website that commonly turns up as the first or second entry in a variety of searches — especially for keywords that are not part of the site's name — can be said to have a lot of Google juice. It is frequently used by bloggers and webmasters. Google's PageRank system plays a large role in ranking results for a given search. It works by counting how many times a page has been linked to and by the "quality" of those links — namely how many times the page that is linking has itself been linked to. In this way, sites with high Google rankings (i.e., lots of "Google juice") can offer to "share" or "give" Google juice to a less popular site. A link from a site with less Google juice can also be helpful in this regard to a lesser degree: there is less juice to give.

Google's 2005 April Fools Day hoax, the fictitious drink Google Gulp, was a take on the term "Google Juice."

Googleating

Googleating pronounced *Google-ating* describes the increasing practise of constructing a website primarily for the purpose of appearing high in the search engine Google.

With the importance of a high Google ranking becoming more apparent, the majority of experienced SEOs (search engine optimizers) know the many tactics used to manipulate the Google algorithm.

A **Googleated** page is often used as a gateway page to the website that the SEO is developing. A small homepage is constructed, heavily optimized for a particular target key-phrase, and from this page hyperlinks are placed to various pages within the main website.

It should be noted that Googleating is not a term used in the SEO industry.

Gateway pages are frowned upon by Google, which means the SEO has to ensure the page is constructed in such a way that the page is relevant and contains an adequate amount of content so as not to be seen as just a gateway page. This leads to a balancing act where the optimizer has to construct a page primarily for the search engines, but designed to look as though it is for the visitor.

In recent months Googleating has taken a new twist, a number of SEOs have been competing in search engine optimization competitions, where the goal is to gain the highest Google placement for a set keyword. This has led to a number of Googleated pages appearing for the phrases Serps and Mangeur de cigogne.

The original serps competition began in January 2004 and ended in April 2004 and the Mangeur de cigogne competition began in March 2004 and ended in June 2004.

Googlebait

Googlebait is jargon for words or phrases that will affect Google searches. These usually involve celebrities and sexual situations that have very little or nothing to do with the article's main point. It is frequently used by bloggers and webmasters to raise their Google PageRank. It is similar to a Google bomb, in that it seeks to influence PageRank, but it is done on the individual page, rather than linking to it several times. Work best with pages with a lot of Google juice. Although it has existed since at least November of 2003, it is more passive than Googlebombing.

SEO contest

In **SEO (search engine optimization) contests**, webmasters compete to rank best on Google for a given (usually nonsense) keyword or keyword combination. They have become an often important method for webmasters to promote their web sites and gain web traffic. While the contestants compete for prizes, fame or glory, the organising body often benefits as well.

History

The *nigritude ultramarine* competition by SearchGuild is widely acclaimed as the mother of all SEO contests. It was started on May 7, 2004 and was won two months later by Anil Dash.

On September 1 of the same year, webmasters were challenged to rank #1 on Google in three months time for the search phrase *seraphim proudleduck*.

In the first quarter of 2005, people were competing for the term *loquine glupe*, spawning web sites ranging from shampoo advertising to holiday resorts. The page that won in the end looked rather boring, and used lots of questionable techniques like "keyword stuffing".

Internationally, in 2005 two major contests took place in Europe. In Germany the Hommingberger Gepardenforelle by the computer magazine c't spawned almost 4 million results. The goal was to find out how search engines rank sites. In Poland almost at same time the Polish SEO community organized the msnbetter thangoogle contest. It topped the 4 million but failed to reach it's goal to promote SEO in Poland and to get search engines companies attention for the polish market. Currently at least one contest is taking place in France.

A competition ran from January 1, 2006 to March 1, 2006 and carried the term redscowl bluesingsky, another set of made-up words. It was sponsored by SEOLogs. Shoemoney won this contest, and since he donated the winner's money, he donated it to the number 2 winner.

A contest that had been announced earlier - but only started on January 15, 2006 - is the one by V7N SEO forum administrator John Scott and another search engine optimizer, WebGuerrilla. In this particular contest, both competitions use the same search phrase *v7ndotcom elursrebmem*, but each has its own set of special rules.

The basics

All these contests appear to be based on a number of common factors:

- In simple words, a SEO contest invites webmasters to trick the search engines. Some webmasters resort to spam, while others use white-hat optimization techniques (like providing good content covering the competition, or optimizing page titles).
- While there are many search engines around, they all seem to focus on Google in particular. Google is known to be a difficult search engine to rank well on, especially for new web sites.
- Most SEO contests expect people to optimize a single web page for a non-existent phrase of two silly words. The main reason for this is to keep existing web sites from getting a head start. But at the same time it makes sure that regular internet searchers won't be bombarded with "spammy" results when searching the web for "regular" information.
- Blogs seem to do well at these challenges, indicating in a way that pages with valuable content are preferred by search engines over regular websites, especially when it comes to newsworthy and fresh information of a temporary nature.

The differences

Certain special rules and limitations are invented to set contest apart from the rest. Often, these limitations will make it harder to benefit from the ranking algorithm - including quirks - of the targeted search engine. For example, the January 2006 Redscowl Bluesingsky contest issued by SEOLogs is open for new domains only. That means that the contestants cannot benefit from the ranking advantage old web sites are thought to have over new ones. An example of that is the age advantage Anil Dash' blog page had over the

well-received but brand new Nigritude Ultramarine FAQ - respectively ended 1st and 6th in the Nigritude Ultramarine challenge. Most likely, the Redscowl Bluesingsky game will be won by a domain of the style redscowl-bluesingsky.com which is bound to attract natural links, and benefit from the fact that the URL is made up entirely of the search words.

Another special rule that fits well with the 'purpose' of SEO contests today is the obligation to 'link back' to the organizing body - often a search engine optimization blog or forum. Since a web document's ranking on major search engines like Yahoo!, Google or MSN Search is mainly determined by internet hyperlinks pointing to that document, forcing webmasters to link to a web site is quite a powerful way to increase its web presence... Good example are the contest announced by V7N and its counterpart by WebGuerrilla. While the first of these originally required the contestants to link to V7N forums, the second forbids its players to do just that. Instead a special link to Google engineer Matt Cutts' blog is imperative. Because of this rivalry, both the rules and prize money on both these SEO contests were updated regularly up until the official start date of January 15, 2006.

About the author

Nicolae Sfetcu

Owner and manager with MultiMedia SRL and MultiMedia Publishing House.

Project Coordinator for European Teleworking Development Romania (ETD)

Member of Rotary Club Bucuresti Atheneum

Cofounder and ex-president of the Mehedinti Branch of Romanian Association for Electronic Industry and Software

Initiator, cofounder and president of Romanian Association for Telework and Teleactivities

Member of Internet Society

Initiator, cofounder and ex-president of Romanian Teleworking Society

Cofounder and ex-president of the Mehedinti Branch of the General Association of Engineers in Romania

Physicist engineer - Bachelor of Physics, Major Nuclear Physics. Master of Philosophy.

Contact

Email: nicolae@sfetcu.com
Online Media: https://www.telework.ro/

Facebook/Messenger: https://www.facebook.com/nicolae.sfetcu
Twitter: http://twitter.com/nicolae
LinkedIn: http://www.linkedin.com/in/nicolaesfetcu
YouTube: https://www.youtube.com/c/NicolaeSfetcu